

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:06:26 ; Search time 59 Seconds
(without alignments)
4635.127 Million cell updates/sec

Title: US-09-442-489F-2

Perfect score: 14575

Sequence: 1 MAAASYDQLKQVEALKMEN.....SSQTQSPKRHSGLVTVS 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Pirl:*

2: Pirl:*

3: Pirl:*

4: Pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	14533	99.7	2843	1	REHUAP		adenomatous polypo
2	13148	90.2	2845	2	I49505		adenomatous polypo
3	3508.5	24.1	2274	2	T30258		adenomatous polypo
4	1778.5	12.2	2416	2	T13825		adenomatous polypo
5	638.5	4.4	5327	2	T13564		microtubule-associ
6	518	3.6	2271	2	F90073		hypothetical prote
7	517	3.5	2232	2	T34434		hypothetical prote
8	516.5	3.5	1186	2	T23327		adenomatous polypo
9	514.5	3.5	1188	2	T23330		hypothetical prote
10	477	3.3	2774	2	A43359		microtubule-associ
11	452	3.1	3507	2	T34513		hypothetical prote
12	442	3.0	4776	2	E95206		cell wall surface
13	428.5	2.9	2722	2	T20532		hypothetical prote
14	426.5	2.9	2738	2	E88320		protein F07A11.6
15	419.5	2.9	1791	2	T02345		hypothetical prote
16	417	2.9	2888	2	I49477		alpha-A-crystallin
17	416	2.9	2938	2	T30249		cell proliferation
18	412.5	2.8	3488	2	T34418		hypothetical prote
19	410	2.8	2187	2	T30826		nascent polypeptid
20	409	2.8	2526	2	T20531		hypothetical prote
21	406	2.8	5170	2	T15348		hypothetical prote
22	405	2.8	3542	2	T42730		Bassoon protein -
23	396	2.7	6713	2	E89921		hypothetical prote
24	395	2.7	3924	2	S37431		ankyrin 2, neurona
25	392.5	2.7	3147	2	T18674		hypothetical prote
26	391	2.7	5105	2	T32650		hypothetical prote
27	387.5	2.7	3187	2	JC5837		364K Golgi complex
28	384	2.6	1459	2	T32271		hypothetical prote
29	382.5	2.6	2176	2	T13806		toucan gene protei

30	381	2.6	2364	2	A56577	microtubule-associ
31	380.5	2.6	4377	2	A55575	ankyrin 3, long sp
32	375.5	2.6	2248	2	A35938	profilagrin - hum
33	375.5	2.6	3562	2	A47171	chondroitin sulfat
34	375	2.6	3968	2	A44265	trithorax homolog
35	373	2.6	971	2	T19431	hypothetical prote
36	373	2.6	3256	2	A48666	cell proliferation
37	372.5	2.6	2253	2	T30336	nuclear/mitotic ap
38	372	2.6	2649	2	T51023	hypothetical prote
39	369.5	2.5	2464	1	QRM9P1	microtubule-associ
40	368.5	2.5	2453	2	S60254	nuclear receptor c
41	367.5	2.5	2346	2	T13829	tpx homolog - fru1
42	365.5	2.5	1560	2	T02885	peroxisome prolife
43	365	2.5	3869	2	A48205	All-1 protein -GTE
44	364	2.5	2441	2	D71623	erythrocyte membra
45	364	2.5	2897	2	B48666	cell proliferation

ALIGNMENTS

RESULT 1

REHUAP

adenomatous polyposis coli protein - human

N;Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1991 #sequence, revision 31-Dec-1991 #text_change 21-Jul-2000

C;Accession: A37261; B39658; A44928; A49319; I54271

R;Kinzie, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith, chul, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.

Science 253, 661-665, 1991

A;Title: Identification of PAP locus genes from chromosome 5q21.

A;Reference number: A37261; MUID:91335210; PMID:1651562

A;Accession: A37261

A;Molecule type: mRNA

A;Residues: 1-2843 <KIN>

A;Cross-references: GB:W74088; NID:G182396; PIDN:AAA03586.1; PID:G182397

R;Joslyn, G.; Carlson, M.; Thlaveris, A.; Alberts, H.; Gelbert, L.; Samowitz, W.; Grode arrington, J.; McPherson, J.; Wasmuth, J.; Le Faslier, D.; Abderrahim, H.; Cohen, D.; Lei Cell 66, 601-613, 1991

A;Title: Identification of deletion mutations and three new genes at the familial polypos

A;Reference number: A39658; MUID:91330307; PMID:1678319

A;Accession: B39658

A;Molecule type: DNA

A;Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P',

A;Cross-references: GB:W73548; NID:G190183; PIDN:AAA0354.1; PID:G190184

R;Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzie, K.W.; Vogelste

Cancer Res. 52, 643-645, 1992

A;Title: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a cc

A;Reference number: A44928; MUID:92119623; PMID:1310068

A;Accession: A44928

A;Molecule type: DNA

A;Residues: 1506-1525 <WIK>

A;Cross-references: GB:S78214; NID:G243541; PIDN:AA21145.1; PID:G243542

A;Note: sequence extracted from NCBI backbone (NCBIN:78214, NCBI:P:78218)

R;Spurio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelbert

Cell 75, 951-957, 1993

A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.

A;Reference number: A49319; MUID:94073973; PMID:8252630

A;Accession: A49319

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 'G', 143-171, 'P', 173-179 <SPI>

A;Cross-references: GB:S67787; NID:G461061; PIDN:AAD13997.1; PID:G4261697

R;Laubertz, S.; Ballhausen, W.G.

Hum. Genet. 90, 650-652, 1993

A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly

A;Reference number: I54271; MUID:93186137; PMID:8383094

A;Accession: I54271

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-4 <LAM>

A;Cross-references: GB:S56365; NID:G256243; PIDN:AAD14918.1; PID:G4262770

1921 QGPXILQKQFTFPOSSKIDIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
 Db QGPXILQKQFTFPOSSKIDIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
 1981 NKNEPIKETEPDPSQGPSPQASGYAPKPFHVEDTPVCFSRNSSLSLSLSDIDEDLLQ 2040
 Db NKNEPIKETEPDPSQGPSPQASGYAPKPFHVEDTPVCFSRNSSLSLSLSDIDEDLLQ 2040
 2041 ECISAMPKPKKPSRLKGNKHSRNNMGILGEDLTLDLKDIOQPSHGLSPDSENF 2100
 Db ECISAMPKPKKPSRLKGNKHSRNNMGILGEDLTLDLKDIOQPSHGLSPDSENF 2100
 2101 WKAIQEGANSIVSSHQAAAAACLSQASSDSDLSLSKSGISLSPHLLTPDQEKPT 2160
 Db WKAIQEGANSIVSSHQAAAAACLSQASSDSDLSLSKSGISLSPHLLTPDQEKPT 2160
 2161 SNKGRIILKPGKSTLETKLIESBKIGKGGKVKYKSLITGKVRNSISISQMKQPLQAN 2220
 Db SNKGRIILKPGKSTLETKLIESBKIGKGGKVKYKSLITGKVRNSISISQMKQPLQAN 2220
 2221 MFSISGRGTMIHIGVRNSSSTSPVSKGPPKPTPASKSPSEGQTATTPRGAKPVS 2280
 Db MFSISGRGTMIHIGVRNSSSTSPVSKGPPKPTPASKSPSEGQTATTPRGAKPVS 2280
 2281 ELSVARQTSQIGSSKAPSRGSDSTPSRPAQPLSRPIQSPGRNSISPGNGISPPN 2340
 Db ELSVARQTSQIGSSKAPSRGSDSTPSRPAQPLSRPIQSPGRNSISPGNGISPPN 2340
 2341 KLSQLPRTSSPTASTKSGSGKMSYTPSGRQMSQNLTKQTLGSKNASSIPRSESAGK 2400
 Db KLSQLPRTSSPTASTKSGSGKMSYTPSGRQMSQNLTKQTLGSKNASSIPRSESAGK 2400
 2401 LNOMNNGANKKVELSRMSTKSGESDRERPVLRQSTFIKEAPSPILRRKLESA 2460
 Db LNOMNNGANKKVELSRMSTKSGESDRERPVLRQSTFIKEAPSPILRRKLESA 2460
 2461 SFESLSPSRPASPSPSOAQTVPVLSPLPDMSLTHSSVQAGWRKLPNLSPTTIEYNDG 2520
 Db SFESLSPSRPASPSPSOAQTVPVLSPLPDMSLTHSSVQAGWRKLPNLSPTTIEYNDG 2520
 2521 RPAKRHDIAARSHSESPRLPINRSGTWKREHSHKSSSLPRVSTWRTTSGSSSILSASSES 2580
 Db RPAKRHDIAARSHSESPRLPINRSGTWKREHSHKSSSLPRVSTWRTTSGSSSILSASSES 2580
 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFPSTNSTQTVSSGATNGAES 2640
 Db SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFPSTNSTQTVSSGATNGAES 2640
 2641 KTLIYQMAVASKTEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKANPNIKSDKN 2700
 Db KTLIYQMAVASKTEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKANPNIKSDKN 2700
 2701 QAKQNVGNGSVPMRTVGLNRLTSPFOVADPOKGTETIKPGQNNPVVSETNESPIVER 2760
 Db QAKQNVGNGSVPMRTVGLNRLTSPFOVADPOKGTETIKPGQNNPVVSETNESPIVER 2760
 2761 PFSSSSSKHSSPSTVAARVTPFNPNPSRKSADSTGARPQIPTPVNNNTKKRDSKT 2820
 Db PFSSSSSKHSSPSTVAARVTPFNPNPSRKSADSTGARPQIPTPVNNNTKKRDSKT 2820
 2821 DSTEESGCTQPKRHSGLVTSV 2843
 Db DSTEESGCTQPKRHSGLVTSV 2843

RESULT 2

149505
 N;Adenomatous polyposis coli protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 C;Accession: I49505

R;Su, L. Science 256, 668-670, 1992
 A;Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the
 A;Reference number: I49505; PMID:92263101; PMID:1350108
 A;Accession: I49505
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-2845 <RES>
 A;Cross-references: GB:M88127; NID:gl91991; PIDN:AB59632.1; PID:gl91992
 C;Superfamily: adenomatous polyposis coli protein

Query Match 90.2%; Score 13148; DB 2; Length 2845;
 Best Local Similarity 90.0%; Pred. No. 0;
 Matches 2566; Conservative 113; Mismatches 159; Indels 12; Gaps 10;

QY 1 MAASVDQLLKQVEALKMENSLRQLEDSNHLTKLETEASNKMKVLOLOSGIDEAM 60
 Db 1 MAASVDQLLKQVEALKMENSLRQLEDSNHLTKLETEASNKMKVLOLOSGIDEAM 60
 QY 61 ASSQIDLLERLXELNLDSSNFPVGLRSKMSLSRSGSREGSVSSRSGECPVPMGSPFR 120
 Db 61 -TSGQIDLLERLKEFNLD-SNFPVGLRSKMSLSRSGSREGSVSSRSGECPVPMGSPFR 118
 QY 121 RGFVNGSRESTGYLELEKEKERSILLADLKEEKEKQWYQAOLNLIKRIIDSLPTENFSL 180
 Db 119 RTFVNGSRESTGYLELEKEKERSILLADLKEEKEKQWYQAOLNLIKRIIDSLPTENFSL 178
 QY 181 QTDLTROLEVEARQIEVAMEEQLGTCODMEKRAQRIARIQIOTEKIDILRIOLLOSOAT 240
 Db 179 QTDLTROLEVEARQIEVAMEEQLGTCODMEKRAQRIARIQIOTEKIDILRIOLLOSOAA 238
 QY 241 EAERSSQKHETGSHDAERQNEGGVGEINNATSGNGQSTTRMDHETAVLSSSSTHSA 300
 Db 239 EAERSSQSRHDAASHEAGRQHEGVAESNTAASSSQSPATRVDTETASVSSSGTHSA 298
 QY 301 PRRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
 Db 299 PRRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 358
 QY 361 HGNDKDSVLLGNRSGSKEARASAAALHNIHSQPDCKRGRREIRVLHLEQRAYCETC 420
 Db 359 HGNDKDSVLLGNRSGSKEARASAAALHNIHSQPDCKRGRREIRVLHLEQRAYCETC 418
 QY 421 WEWOEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRHANNELGLOIAELLQ 480
 Db 419 WEWOEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRHANNELGLOIAELLQ 478
 QY 481 VDCWYGLTNDHYGITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSESED 540
 Db 479 VDCWYGLTNDHYGITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSESED 538
 QY 541 QQVIASVLRNLWRADVNSKKTLEVGSKVKALECELEVKKESTLKVLSALMNLSAHCT 600
 Db 539 QQVIASVLRNLWRADVNSKKTLEVGSKVKALECELEVKKESTLKVLSALMNLSAHCT 598
 QY 601 ENKADI CAVDGALAFVLTLYRSQTNLTALIESGGGILRVNSSLIATNEDHROILRENN 660
 Db 599 ENKADICAVDGAFLVLTLYRSQTNLTALIESGGGILRVNSSLIATNEDHROILRENN 658
 QY 661 CIQTLLOHLKSHSLTIVSNACGTLNLSARNPKQDEALWDMGVSMKLNLIHSHKHMAM 720
 Db 659 CIQTLLOHLKSHSLTIVSNACGTLNLSARNPKQDEALWDMGVSMKLNLIHSHKHMAM 718
 QY 721 GSAAALRLMNRPAKYKDNIMSPGSSLPKSHYRKOKALEAELDAQHLSSTFDNIDNLS 780
 Db 719 GSAAALRLMNRPAKYKDNIMSPGSSLPKSHYRKOKALEAELDAQHLSSTFDNIDNLS 778
 QY 781 PKASHRSKORHOSLYGDYVDFDTNRHDDNRSDNFTNGMTVLSPLYNTTVPSPSSSSRGS 840
 Db 779 PKASHRSKORHONLYGDYAFDANRHHDSRSDNFTNGMTVLSPLYNTTVPSPSSSSRGS 838
 QY 841 LDSSSEKDRSLERERIGLGNYPHYPATNPGTSSKRGLOISTTAAQIAKYMEEVSAIHTS 900

Db 839 LDSSRSKDRSLERBERGICLSAYHTTENVAGTSSKRGIGIITTTAAQIAKVMEEVSAIHTS 898
Qy 901 QEDRSSGSTTEHLCVTDERNALRRSSAAHTSNVNTFTKSENSNRTCSMPYAKLEYKRSS 960
Db 899 QDRSSASTTEPHCVADRSAAARSASHTSNVNTFTKSENSNRTCSMPYAKVEYKRSS 958
Qy 961 NDLSNVSNDGCKRGQWKPSIESYSDDESCKFCYGYPADLAHKIHSANHMDDNDGE 1020
Db 959 NDLSNVSNTSSGCKRGQWKPSIESYSDDESCKFCYGYPADLAHKIHSANHMDDNDGE 1018
Qy 1021 LDTPINYLKYSDEOLNSGRSPQNERWARPKHIIEDIKQSPQRQSRNQSTTPYVTE 1080
Db 1019 LDTPINYLKYSDEOLNSGRSPQNERWARPKHIIEDIKQSPQRQSRNQSTTPYVTE 1078
Qy 1081 STDDKHLKFQHFQGOECVPSYRGANGSETNVRGSHNGINQVNSQSLCOEDDYEDDKP 1140
Db 1079 NTDDKHLKFQHFQGOECVPSYRGANGSETNVRGSHNGINQVNSQSLCOEDDYEDDKP 1138
Qy 1141 TNSERYSEEQH-EEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFYSK 1199
Db 1139 TNSERYSEEQH-EEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFYSK 1198
Qy 1200 SSSQSSKTEHWSSESNTSPSSNAKRONLHPSSAQSGQKQKATCKVSSINQETI 1259
Db 1199 NSSAQSTKPEHLSSESNTAVPPSSNAKRONLHPSSAQSGQKQKATCKVSSINQETI 1257
Qy 1260 QTYCVEDTPICFSRCSLSLSSAEDBICGNQTTQEADSANTLQIABIKGKITRGAEDP 1319
Db 1258 QTYCVEDTPICFSRCSLSLSSAEDBICGNQTTQEADSANTLQIABIKGKITRGAEDP 1317
Qy 1320 VSEVPVAVSQHPTKSRKLOGSSLSSESARH-KAVEFPSSGAKGPKSGAQTPKPPPEHYVQ 1378
Db 1318 ATEVPVAVSQHPTKSRKLOGSSLSSESARH-KAVEFPSSGAKGPKSGAQTPKPPPEHYVQ 1377
Qy 1379 ETPLMFRCTSVSLDSEFERSIASSVQSPBCSMVSGIISPDLDPSPQCTMPPRSKRT 1438
Db 1378 ETPLMFRCTSVSLDSEFERSIASSVQSPBCSMVSGIISPDLDPSPQCTMPPRSKRT 1437
Qy 1439 PPPPQTAQTKREVPKNAKPAEKRESGPKQAAVNAAVQVQLPADTLHLHFEATSTPD 1498
Db 1438 PPPPQTAQTKREVPKNAKPAEKRESGPKQAAVNAAVQVQLPADTLHLHFEATSTPD 1497
Qy 1499 GFSCSSLSLSALSDLEPIQKDVLRINPPVQENDNGNETESEPESKSNENQEKAKTID 1558
Db 1498 GFSCSSLSLSALSDLEPIQKDVLRINPPVQENDNGNETESEPESKSNENQEKAKTID 1556
Qy 1559 SEKDLDDSDDDDIIELEECIIISAMPTKSRKGGKPAQATASKLPPPVARKPSQLPVYKLL 1618
Db 1557 SEKDLDDSDDDDIIELEECIIISAMPTKSRKGGKPAQATASKLPPPVARKPSQLPVYKLL 1616
Qy 1619 PSQNLQPKHVSFTPGDDPRVYCVGEGTPIINSTATSLDITIESEPPHLAGSVRG 1678
Db 1617 PAQNLQPKHVSFTPGDDPRVYCVGEGTPIINSTATSLDITIESEPPHLAGSVRG 1676
Qy 1679 AQGEFEKRTIPIEGRSTDEAOGKTSVTIPELDNKAEEGDIIAECINSAMPKGS 1738
Db 1677 IQGEFEKRTIPIEGRSTDEAOGKTSVTIPELDNKAEEGDIIAECINSAMPKGS 1736
Qy 1739 KPRVKIMQVQQAASSAPKNQLDGKKKPTSPVKPIPONTYRTRVRKADSKN 1798
Db 1737 KPRVKIMQVQQAASSAPKNQLDGKKKPTSPVKPIPONTYRTRVRKADSKN 1796
Qy 1799 LNAERFSDNKKQKQNLKNSXDFNDKLPNNEDVRGSPAPDSPPHYTPIGTPVCFSR 1858
Db 1797 VNTTEFSDNKKQKQNLKNSXDFNDKLPNNEDVRGSPAPDSPPHYTPIGTPVCFSR 1856
Qy 1859 NDLSLSDFDDDVDVLSREKAEJLRKAKENKESPAKVTSHTELTSNQOANKTCAIAKQPI 1918
Db 1857 NDLSLSDFDDDVDVLSREKAEJLRKAKENKESPAKVTSHTELTSNQOANKTCAIAKQPI 1916
Qy 1919 NRQOKPILQKSTFPQSSKIDPRGAATDEKLQNEFAIENTPVCFSHNSLSLSDIDOE 1978
Db 1917 NRAQSPVLQKSTFPQSSKIDPRGAATDEKLQNEFAIENTPVCFSHNSLSLSDIDOE 1976

Qy 1979 -NNKNEPIKETEPDPSQGEPSKPAQAGYAPKSFHVEDTPVCFSRNSLSLSIDSEDD 2037
Db 1977 NNNKNESEPIKEAPEANQGEPSKPAQAGYAPKSFHVEDTPVCFSRNSLSLSIDSEDD 2036
Qy 2038 LIQECISSAMPKKKPSRLKGNKHSERNMGIGILGEDLTLDLKOIQRPDSEHGLSPQSE 2097
Db 2037 LIQECISSAMPKKKPSRLKSEKQSPKVGIGILAEDLTLDLKOIQRPDSEHGLSPQSE 2096
Qy 2098 NFDWKAIOGANSIVSSLSHQ-AAAAACLSRQASSDSILSKSGISLCSGSPHLPDQEE 2156
Db 2097 NFDWKAIOGANSIVSSLSHQ-AAAAACLSRQASSDSILSKSGISLCSGSPHLPDQEE 2156
Qy 2157 KPFTSNKGRPIILKPGKEKSTLETKTIESKSGIKGKGVYKSLITKVRNSBEISQMKQP 2216
Db 2157 KPFTSNKGRPIILKPGKEKSTLEAKTIESKSGIKGKGVYKSLITKVRNSBEISQMKQP 2216
Qy 2217 LOANMPSISRGRTHIIPGVNRSSSTSPVSKKGPPKLTASKSPSEQOTATTSPRGAKP 2276
Db 2217 LFTNPSISRGRTHIIPGLANSSSTSPVSKKGPPKLTASKSPSEQOTATTSPRGAKP 2276
Qy 2277 SVKSELSPVAROTSOIGGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISPGRNGI 2336
Db 2277 AGKSELSPITROTQISGNSKSGSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISPGRNGI 2336
Qy 2337 SPNKLSQLPRTSSPSTASTKSSGSKMSYTSRQMSQOQLTKOTGLSKVASSIPRSES 2396
Db 2337 SPNKLSQLPRTSSPSTASTKSSGSKMSYTSRQMSQOQLTKOTGLSKVASSIPRSES 2396
Qy 2397 ASKGLNQMNGNKGANKVELSRMSTKSSGESDRSERPVLVROSTFIKEAPSTLRKL 2456
Db 2397 ASKGLNQMNGNKGANKVELSRMSTKSSGESDRSERPVLVROSTFIKEAPSTLRKL 2456
Qy 2457 BESASFESLSPSSRPASPTRSOACTPVLSPSLPMSLSTHSSVQAGGWRKLPNLSPTIE 2516
Db 2457 BESASFESLSPSSRPASPTRSOACTPVLSPSLPMSLSTHSSVQAGGWRKLPNLSPTIE 2516
Qy 2517 YNDGRPAKRHDIAKSHSPSRPLPINRGTWKRSHKSHSSLPVSTWRTGSSSSILSA 2576
Db 2517 YNDGRPAKRHDIAKSHSPSRPLPINRGTWKRSHKSHSSLPVSTWRTGSSSSILSA 2576
Qy 2577 SSESSEKAKSEDEKXVANSISGTQSKENQVSAKTWKIKENEFSPNTNSTQTVSSGATN 2636
Db 2577 SSESSEKAKSEDEKXVANSISGTQSKENQVSAKTWKIKENEFSPNTNSTQTVSSGATN 2636
Qy 2637 GAESKTLIYQMAPAVSKTDEYVWRIEDCPINNPSRSGPTNTPPVIDSVSEKANPNKD 2696
Db 2637 GAESKTLIYQMAPAVSKTDEYVWRIEDCPINNPSRSGPTNTPPVIDSVSEKANPNKD 2696
Qy 2697 SKDNQ---AKONVNGSVPMRTVGLNRLTSFIQVADPOKGTGTEIKGQNNPVPVSENE 2753
Db 2697 SKDNQTHGKQSVSGS-PVQTVGLETRLNLSFVQEAPEQKGTGTEIKGQNNPVPVSENE 2755
Qy 2754 SPIVERTPFPSSSSSKHSPSGTVAARVTPNPNPSRKSADSTARPSCQIPPPVNNNT 2813
Db 2756 TCIAERTFPFPSSSSSKHSPSGTVAARVTPNPNPSRKSADSTARPSCQIPPPVNNNT 2815
Qy 2814 KKRDSKTDSTESSGTQPKRHSGLYVTSV 2843
Db 2816 KKRDSKTDITESSGAQPKRHSGLYVTSV 2845

RESULT 3

T30258

adenomatous polyposis coli protein 2 - mouse

N:Alternate names: APC2 protein

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000

C:Accession: T30258

R:van Es, J.H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Myles, A.; Kuipers, J.

Curr. Biol. 9, 105-108, 1999

A:Title: Identification of APC2, a homologue of the adenomatous polyposis coli tumour sui

A:Reference number: 220796; MUID:99147086; PMID:10021369

[illegible]

QY	875	KGLOI	STTAQIAK	WMEV	SAIH	TSQED	RSGS	ITLH	CHV	TDE	NAL	RRSSA	HTHNT	934
		:::												
DB	834	KAKAL	ALAVARI	DL	VEDI	SA	LH	TSSD	DSLSS	----	GD	QOEAP	REGRAQSC	887
		:::												
QY	935	YNF	T--	KSENS	NRT	CS	MP	VAK	LEY	KRS	ND	SLNS	VSS	992
DB	888	RTG	EGRR	EAGS	RA	HP	LL	RL	KA	HT	SL	ND	SLNS	936
QY	993	KFC	SGVQ	VP	AD	LA	HK	I	GA	NH	MD	ND	GEL	1052
DB	937	--	CPLAAL	----	AHRD	----								
QY	1053	KH	I	E	I	E	I	K	Q	S	R	O	R	1112
DB	957	R	I	D	L	D	L	P	S	R	A	L	P	1002
QY	1113	N	R	V	G	N	H	I	Q	N	V	S	O	1172
DB	1003	----	GVR	P	V	G	P	G	T	----				1011
QY	1173	V	Q	P	I	D	I	S	L	K	Y	A	T	1232
DB	1014	----	PG	A	K	Q	A	W	----					1021
QY	1233	P	S	A	Q	S	R	G	O	P	A	K	T	1292
DB	1022	-	I	P	A	D	S	L	K	V	P	E	L	1077
QY	1293	T	Q	A	D	S	A	N	T	L	Q	A	B	1343
DB	1078	A	E	N	L	D	S	D	S	L	E	E	A	1128
QY	1344	S	E	S	A	R	H	K	A	V	E	F	P	1403
DB	1129	V	E	D	A	----	TP	S	S	S	E	N	C	1168
QY	1404	S	V	S	E	P	C	S	G	V	G	I	I	1457
DB	1169	S	I	P	S	D	F	C	S	L	G	T	V	1228
QY	1458	P	T	A	K	E	S	E	G	P	K	A	V	1517
DB	1229	D	I	A	D	C	H	E	R	O	P	P	E	1276
QY	1518	K	D	V	E	L	A	T	M	P	P	V	----	1573
						</								

770 -----ENATKTSY----- 778
1110 SETNRVGNHGINVNSQLCQEDDYEDDKPTNYSERYSEREEBEERPTNYGIKNEE 1169
779 -----QETDL--DQPTDSLVAE-----NQIESDLDISGPAGG 811
1170 KRHVDPIDYSLKVATDIP--SQKQSFSSKSSQGSKTH-----MSSSENTSTPSS 1223
812 KSTITPP-----AETVPESEGEIILLIILDDSVKCYQEDTTPYVISAASVTDLRVAA 864
1224 NAKRONQLHPS--SAQSRSGOPO--KAATCKVSSINOETIQTVCVEDTPIPCRSKSSLS 1278
865 KADRAEYKPEVREVTSEKGA PKLPKLSQCGSGSYTEPEKINCEGTPTGFRYDLSL 924
1279 SLSSAEDRIG--CNQOTQFADSANTLQIAEIKGIGKTRSADPVPVPAVSHQPTKSSRL 1337
925 SL-----DESGKANQAIUGTD-----ADIKPKLEKQEBQ----- 954
1338 QGSSLSSESARHKAVERFPSSGAKSPSKGAQTPKPPPHVVOETPLMBSRCTSVSSLSDFE 1397
955 -----SOPAEQVLTKPPTQANS-----ALETPLMFSRRSSMDSLVHP 992
1398 SRSLA-----SSVQSEPCSGMVSGIISPSDLDPDPSGQTMPPSRSKTTPPPPTQATKREV 1452
993 DVDVANCDDKSSVSD--FSRLASGVISFSEIPDSTQSMPOS----- 1033
1453 PKNKAPTAKRESGPKQAANVAORVQVLPDADTLHFATSTPDGSCSSLSALSLD 1512
1034 PRNSVAGSGQNDVPPVVPVIFASLOPLRSVFE--DDLSSFNVEHTPAQFSTATISLSLI- 1091
1513 BPFIQKDYELRIMPVQENDGNE-----TESEQPKESNENQKE--AEKTIID 1558
1092 -----VDDEKAPAVWTEDEDELALLANCINMGQRKPTAVKSTVNSVDVAETIR 1144
1559 S-----EKD-----LLDD--SDDDI----- 1572
1145 SYCTEDTPALLSKVPNTNLSVISMSSTDPKDATAGQAQMAHQLSDDVSSNASDCGAS 1204
1573 --EILECIISAMPTKSRGKGAATASKLPPPVARK--PSQLPVYKLLPSONR--QPOKHV 1630
1205 GHLLQOCIRDM-----KKPLGEATSDPTAMLRGNGELFGY--LPS----- 1244
1631 SPTGDDMPRVYCVGTPTNFSTATSLDLATIES-----PPNELAAGEVGRGAOSGEFEK 1686
1245 ---ADEMNK--FLVEDSPCNFVSVGLSNLTVGSSLVGPVQLKETE--PSSADQNPENKR 1298
1687 RDTITEGRSTDEAQQGKTSVTIPELDNKAEBGDILAC----- 1727
1299 SLANKSKRRPPHWQDDSLSSLSIDSEDDTNLLSQAIAGCNRPKSNLGFSSNGKSSSL 1358
1728 -----INSAMPKKGSKHPPRVKIMQVQOASASSAPKNQO-----LDG--KK 1769
1359 SSSQPIANAATSASLSAMTVRSQQOESYSSVSDSDNDNQSKSLFELCILKGMVKT 1418
1770 KKP-----TSVPKPIPONTTEYRT-----RVKKNADSKNLNAER 1803
1419 KEFGARAQOMQOPIVGSSSVQSNPSLQFSLPVLQPSGGQVKQRHHHHHHHRRER 1478
1804 VESDNKOSK--KQNLKXN--SKDFNDKLPNNEDRVKGSFAFDSPHYTPIEGTPTCYFSRND 1860
1479 -----ERKDEKLQECINTGILSKIN--AVPKV--LATSAALEPCH--PMAAT-----T 1523
1861 SLSSLDFFDDDDVDLREKAEALRKAKENKESAEKVTSHTELTNSQOASANKTOA--IAKQPI- 1918
1524 SASALSTAAPDVE---QKAH---ATSNPQQOQSTTHPSSHILNPIDATATVDTARSAA 1577
1919 --NRGQPKPILOKQSTFPQSKDIPRGAATDEKLQNFALN-----TPVCF----- 1963
1578 PNQGNAGNSQNGLET--ATGSKOLDSEDRSSDESQSFIMETVRLDLSALNETCISGASEK 1636
1964 -----SHNSLSLSLSDIOENNNKNEPIKETEPDPS 1995
1637 HKDPDLMLKSLVERLTWTEFVTSAEQLRSSSHNHSS-----NSHKKNSSNNTWNSTCFND 1691

RESULT 5

T13564

microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)

N:Alternate names: hypothetical protein EG:49E4.1

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

R:Accession: T13564

R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.

submitted to the EMBL Data Library, April 1999

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: 217689

1996 QGEPSPQASGVAPKSFHVEDTPVCFSRNSSLSLSID---SEDDLLQECISAMPKKKK 2052
1692 VSPFSVSQTA-----PV-----LASTSLDEDATEARSLHELIEITPTNEQG 1732
2053 PSLKGDNEK-----HSPRNMGILGEDLTLDLKDIOQPSEHGLSPDSENFOWKATCEGA 2108
1733 PESLEGETDTLVNGHADSYSGSGGLNFOL---GGQVQAGVRLEFORLLFNFTSASIMT 1789
2109 NSTVSSIHQAAAAACLSRQASDSD-----SILSL-----KSGI--SIGSPF--HLTP 2152
1790 NSTMIAREALAEALLQPAATDDDTTETFSILNSLDLNIIRPPSGMESLNSCYQDHSOP 1849
2153 DQEEKPFTSNKGPRIILKPEKSTLETYKIESKGIKGGKVKYKSLITKVRANSISGQ 2212
1850 SSRQAMPS--KSPRFARKMFPANLVARALHLAG-----SAESVNSSCNLLDN 1897
2213 MKOP-----LQANMPISIRGRTHIHIPGVNRNSSSTSPVSKKGPPL----- 2253
1898 IKPFLMDELDDSMIVSDISQSEVADGEQDCSWATTISVSNYETAACDDQTMVLOSCFD 1957
2254 -----KTPASKSPSEGQTATTSPRGAKPSVKSEL-----SPVAR 2287
1958 EDEDATNDYSSAESTPKHGSTPSPNRRSLTPKDKERLTAKDRPKTYTIATSCMEAPEAN 2017
2288 QTSQI-----CGSSKAPSRSGRSDTSPRPAQQLSRPI----- 2321
2018 ETLLQELIVEAANVPATSPRANGRRGSAERYKTQIECPALIQPDOPDCPSEQLSIR 2077
2322 -----QSP--GRNSISPRNG-----ISPPNKUS 2343
2078 AMWQQTFTIDINIGHSQETCESTDHPDAGESPECQDQNSETESCDQEPDHLPPPPSIV 2137
2344 QL-----PTSSPSTASTKSSSGKMSYTSRQMSQQNLTKQTGLSKNASSIPRSESA 2397
2138 DLRTSVVKPTTLEPAFAVKLVGRKKPVPVSPYMSQSRN----- 2177
2398 SKGLNOMNNGANKKVELSRMSTKS-----SGESDSRSEPV-----LVRQSTP. 2443
2178 -----SNNAAASKKLTLPSTAKSLVPGSGVRLPAKKXTPTPEPAPALERQGT 2230
2444 IKAPSTLRLKLEASFEPSRSPAPPTSQOATPVLSPLDMSLSTHSSVQAGG 2503
2231 VKD-----EPTNSNVQVPVVE--TKPAQTSPTHRA----- 2258
2504 WRKLPPNLSPTIEVNDGRPAKRDIARSHSESPSLPINRSGTWKREHSHSSSLPRVST 2563
2259 -SKLPTKKGTA---SGSPSK-----AGSPKRIPLAPA---RMTPOEANTSLALAA 2303
2564 WRRTGSSSILSASSESEKAKSEDEKHVNSISGTKQSKENQVSAK--GTWRKIKENEFS 2621
2304 GKSPAASRVVSGRVSTTPPSRNSNLNGSSAAAAAKINQAQSRIANIKRWVDEAKTK 2363
2622 PTNS---TSQTVSSGATNGAESKTLIQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGN 2678
2364 QSSSNLTKQTKSSNMLNANGTKPTLLR-----SSTPD-----N 2397
2679 TPVIDSVSEK 2689
2398 TPSTAGGVKSK 2408

A:Accession: T13564
 A:Status: Preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-5327 <SPA>
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0025392
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A:Note: EG:49E4.1
 C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.4%; Score 638.5; DB 2; Length 5327;
 Best Local Similarity 18.6%; Pred. No. 6.2e-18;
 Matches 610; Conservative 504; Mismatches 1253; Indels 921; Gaps 129;

QY 2 AAASVDQLLKQVEA---LKMENSLRQELE--DNSNHLTKLETSANMKEVLKQLQGS 55
 DB 941 ABEIEIAIAKVEAEAKLEGASAKQSELDVPEQSKIAEVDIIATADIAKSR 1000
 QY 56 EDEAMASSGQIDLLERLKLNLDSNPPGVKLRSKMS-----LRSYG 97
 DB 1001 TEEQLAKPAEEL-----SSPTPEEKLKSTSDTKDQIGAPVDVLPVNLQESLP 1050
 QY 98 SREGSVSRSG-----FCSVPVPMGSPFRPFVNGSRSTGYLBELEKERSLLIAD 149
 DB 1051 BEKFSATIESGATTAPLPEDERIPDQIKEDLVI---BEKYVKEETKEAEAIIVVATVQ 1106
 QY 150 -----KEEKEDWYVAQLNLTKRIDS---LPLTENPFLQDILTRQ 188
 DB 1107 TLPEAAPLAIDTILASATKAPD--ANAEALGELPDSGERVLPKMTFPAQONLLRDV 1163
 QY 189 LEY--EARQTRVAMEOLGTCQMEKKAQRRIQIQIEKILIRIQLLOQSQATEAERS 246
 DB 1164 IKTPDEVADLPVHEADLGL---YEKDSQDANAK-----SISHKEESA 1203
 QY 247 QNKHETGSHDAERNEQGVGETINMATSNGNGSGSTTRMDHETASVLSSTSHSAPRR--- 303
 DB 1204 KEKET--DDEKENK---VGEELGDEPN-----KVD--ISHVLLKESVQBEAKVVV 1249
 QY 304 LTHSLGPKVEMVYSLLSMLGTHDKDMSRTLLAMSSQDSISMROSGLPLLIQLHGN 363
 DB 1250 IETTVEKKQBEIVAEATTVITQENQED---LMEQVKDKKEHEQKIBSGIIT--- 1296
 QY 364 DKOSVLLGNRSGKEAPARASAAHLLIHSQPDQKGRREIRVHLLQIRAYCETCWEW 423
 DB 1297 -----EKEAKKGAST-----PEEK----- 1310
 QY 424 QEAHEPGMDQDNFMPAPVEHQICPAVCLMLKLSFDEEHRHAMNELGLQA---IAELIQ 480
 DB 1311 -ETSDITSDBE--LPA---QLADPTVPKPSAKDRE-----DTGSIESPTIEEAIE 1356
 QY 481 VCEMYGLTNDHYSITLRRVAGMALTNLPFGDVANKATLCSMKGMKRALVAQLKSESDL 540
 DB 1357 VE-----VQAKQEAQ-- 1366
 QY 541 QQVIASVLRNLNLRADVNSKTLREVGSVKALMECALEVK--KESTLKSVLNLSALNLSAHC 599
 DB 1367 KPVPAPBEAIKTEKSPLASKETSRPSATGSKVKEDETKTSKSPYPS----- 1414
 QY 600 TENKADICAVDGAFLVGLTTRVSTQNTLAIIESGGILRVNSVLIATNEHRQI---L 656
 DB 1415 ---RPSEAKDKKSPFASGEAS--RPESVAESVDEAGKAESRRSREKATKDKDESSLDKAK 1470
 QY 657 RENNCLQTLQHLKSLTIVSNACGLTNLNSARNPKDOEALWDMGAVSMLNLIHSKHK 716
 DB 1471 EQSRRESLAESTKPSGIDKESA-----LAKSEASRPESVID-----KSKPSPRRE 1517
 QY 717 MIANG--SAAALRNLMANRPAYKANDIMGPSLSLHYRKQKALFAELDAQHLSSETFN 775
 DB 1518 SIASLKAESTKDEKGAPEK---EASRPGSVVESVKDETEKSKEP-----SRRESI 1566
 QY 776 IDNLSPKASHRSKQRHKQSLGYGVFDTRHNDNRSDNFTNGMTVLSPYLTNTVLPGSS 835

DB 1567 ABSAKPIEPREVSRPESVI--DGIKDSEAKPESRRD-----SP-----LASKE 1608
 QY 836 SSR--GSLDSSRSEKORSLERGERGILGNHYPA-----TENPFTSKRGLOISTAAQIA 888
 DB 1609 ASRPESVLESVKDEFIKSTERSRESVAESFKADSTKDEKSPITSK-----DISPESAVE 1664
 QY 889 KMYEVSALHTSQRSSSGSTTELHCVTDERNARRSAAHTHTNTYNTFKSESNSTCS 948
 DB 1665 NVM-DAPKETSRPESAVGS-----MKDESMKEPSPRRSVKDKAAQASRETS 1710
 QY 949 MPYAKLEYKRSNDLSNLSVS---SNDGYGKRGQWKPSIESYSEDEDEKFCGSGYGPADLA 1005
 DB 1711 RPASVAESAKDGADLKLSRPESTTQSKEAGSIKDEKSPASEEASRPASVAESVKDEA 1770
 QY 1006 HKIHSANHMD--NDGELDTPLNYSLYKSYDEQLNS-----GROSFSQWERW 1049
 DB 1771 EKSKEESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEESRRESVAESKPLPSKEA 1830
 QY 1050 ARPKHI---IEDEIKQSEQ---RQSRNQSTTYP-----VYTESTDDKHLKFPQPHFG 1095
 DB 1831 SRPASVAESIKDEAKSKEESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEE--SR 1889
 QY 1096 QECV-----SPYRSRGANGSEINRVGSHNQNVSQSLCQED---DYEDDKPTNYS 1145
 DB 1890 RESVAESKPLPSKEASRPASVAESIKDEAKSKEESRRESVAESKPLPSKEASRPASVAE 1949
 QY 1146 RYSBE-EQHEEERPTNYSIKYNEEKRHVDIPIDYSLKYATDIPSSQKQSFSSKSSGQ 1204
 DB 1950 SIKDEAKSKEESRRESVAESKPLPSKEASRP-----ASVAESIKDEAKSKEESRR 2001
 QY 1205 SKTEHMSSTSTSTPSSNA-----KRONLHPSSAQSRGSGQPKAATCKVS----- 1252
 DB 2002 ESWAKSPLPSKEASRPASVAESIKDEAKSKEESRRESVAESKPLPSKEASRPASVAES 2061
 QY 1253 ---SINQTIOTYCVEDTPICFSRCSLSSS--SAEDEIGCNQTTQEDS--ANTLOI 1304
 DB 2062 IKDEAKSKEESRRESVAESKPLPSKEASRPASVAESVKDEADKSEESRRESVAESGKA 2121
 QY 1305 AEIKGKIG-----TRSAEDPVSEVPVSOHPKTSRRLQSGLSSESAHKAHKAPEFPG 1357
 DB 2122 QSIKGDOSPLKEVSRPESVAESVKDDPVKSEPRRES--VAGSVTADARD----- 2172
 QY 1358 AKSPKS--GACTPKSPPEHYVQETPLMFSTCTSVSSLDSPESRSIASSVQSEPCSGMVSG 1416
 DB 2173 -QSPLESKGRSPESVDSVKDEAKSKEESRRES-----KTES 2208
 QY 1417 IISPSDLPD--SPGQTMPP--SRSKT-----PPPPPTQATQKRE-----VPKN--KAP 1458
 DB 2209 VIPPKAKDKSPKEVLQPVSMVTETIREDAQPMKPSQAESRRRESIAESIKASSPRDEKSP 2268
 QY 1459 TAEKESGPKQAA--VNAAVQVRVQVLPDADTLHLFATESTPDGFCSSSL----- 1507
 DB 2269 LASKEASRPGSVAESIKYDLDKPOI--IKDKXSTHRESLEDKSAVTSEKSVSRPLSVAS 2328
 QY 1508 -----ALSID-- 1512
 DB 2329 DHEAAVAIEDDAKSSI SPKDKSRPGFVAETVSSPIEATMEFSKIEVVEKSSLSLSIQGG 2388
 QY 1513 -----EPFTQKQVELRIM-----PP 1527
 DB 2389 SGGKLOTSSPVDVAEGDFSHAVASVSTVPTTLTKPAELAQIGAATVSSPLDEALRTPS 2448
 QY 1528 VQENDNGEYTESQPKESNENQKEAKSTIDSEKDLDDDDDDDDDDDDDDDDDDDDDDDDDD 1587
 DB 2449 APEHISADSPAEACASEIASQDSQVTLKESRPAPVAESKDDAAQLKSSVEDLSPVA 2508
 QY 1588 SRKGGKPAQATASKLPPPVPAKPSOLPVYKL-----LPSQNRLOPQKHVSF 1632
 DB 2509 STEISRPASAGETASSPIEAPKDFAEFOAEKAVPLTLIELKGNLTPTLSSPVDVAHASV 2568
 QY 1633 TPGD---DMPRVYCVBGTPIFNSTATSLDLTIESPPNELAAGEGVRGGAQSGEPEKRD 1688
 DB 2569 QPAELSKVDIEK---TASSPIDEAPKSLIGSPAERPE---SPAESAKDAESVEKSKDA 2622

Query Match		3.5%; Score 517; DB 2; Length 2232;
Best Local Similarity		20.4%; Pred. No. 2e-13;
Matches 387; Conservative 247; Mismatches 811; Indels 454; Gaps 69;		
QY	1185	TDIPSSQKQSFs-----F8KSSGGGSSKTEHMSSENSTFSPSNAKRONQLHPSSAQSRs 1240
Db	360	TVVPGSSSTFASSTPIASSSPGTVTAVPGSSSTYSGSTPSAGSSSGTMTSGSTGS 419
QY	1241	GOPOKAATKVSSINQETIQIYCVEDTPICSRc--SSLSLSAEDBIGCNOQTQADS 1298
Db	420	-----TVVAVSSSTFGS-----STPIASSSSSGSTTVVSGSSSTYSGSTPSAGSSs 468
QY	1299	ANTLQIAIKIGKIGTRSAEDPVSEPAVSQHPRTKSSRLQg--SSLSSESARHKAVERPPs 1356
Db	459	AGT--ASTISGSGTATVPGSS--SSVSGSTQASPPSGTMTSGPTGTVTV-VPG 524
QY	1357	CAKSPKSGAQTPKPPPHYQETPLMFSCR-----TSVSLDSFESRIASSVQSE 1408
Db	525	SSTSPAPSSSNPSSPASTGSTIFIGSSSIIIVTSGSTVSGSTGTSQSTLASSTATP 584
QY	1409	PCSGMVSGIIGSPDLDPDP-----GOTMPPSRSKTTPPPPTQAKREVPKNKAPTAKRE 1464
Db	595	GSSSTVPSSSPQSSQAPNWTGSTTFSQTSQSPSPSMN-----PSSSTTSGSSQS 637
QY	1465	-----SGPKQA-----AVNAAVQVQVLPDADTLHLHFATSTPDGFCSSSLGALS 1510
Db	638	TITPEGSTASGPTGSGTSTFVATEVTQSVTPSGSSLGTSQSTNSPSPSSLSPTSGMS 697
QY	1511	--LDHPPFQKQVELRIMPVQENDNGNE-----TESEQPKENENQEKAE----- 1554
Db	698	TITSEP-----SPSSTQSGAGQAGTITTPFPNPSQTSLSLESTGATTSAGSAG 746
QY	1555	--KTIDSEKDLDDDDDDIEILEBECIISAMPTKSRKKGKPAQPAKLPFPVARKPSQL 1612
Db	747	TTMTSPSSSSVGGSSQG-----STPAASTTSGEMTSQGSTQTPGSSSVSTSAAI 795
QY	1613	PVYKLLPQNRLQPKHVSFTPCDDMPRVYVEGTPINFSTAT-----SLSDLTIESPP 1666
Db	796	-----LTJQQSVSTK-----SPGSTVTRPSTVSGSTVSGSTVTVGSTEASTSGSSVASSS 846
QY	1667	NELAAEGVRGGAQSGE-----FEKEDTITPEGRST-DEAQQGKTSVVIPE----- 1712
Db	847	PAPSTSONENPSTSGSSMITQSPYPSQSTSPVESSTTPSPSGPTTLTSTSPSQSTT 906
QY	1713	LDNKAEGDILAEICINAMPKSHKHPRVKIMQVQOASASSAPNKNQLGKKKKP 1772
Db	907	IGSTQGSTPGISITSEEMTSQGSTQTPGSTVTPSTVSDSTSGSTVTV-GSTEGS 965
QY	1773	TSPVKPIPONTETVRVRKNADSKNNAERVFSDNKDKKONLKNKSKDFNDKLNNED 1832
Db	966	SPIPSTQNTNPST-----SSGSSMTQTPQSSQSTSPVESSTSGA----- 1007
QY	1833	RVRGSFAFDPSPHYTPIEGTPTCYCFGRNDSLGLDFDDDDVLSREKAE LRKAKENKESEA 1892
Db	1008	-----TSSSGSP----- 1014
QY	1893	KVTSHELTSNCOOSANKTQAIKQPINRGQPKPILOKOSTPPOSSKDIPIDRGAATDEKLO 1952
Db	1015	-----GTLTISIPSPSPSTIGS-----SQGSTSPV-----STISQGSTETPGSTGVTK-- 1062
QY	1953	NPAIENTPVCFSHNSLSLSLDIOE-----NNKENEPIKETETPPDQCEPSKPOAS 2005
Db	1063	-----PSTVSGSASSGSTATMGTEASSTSGGSSSTPNPFSQTSPTS-GATSSPGSS 1114
QY	2006	GYAPKSFH---VEDTFVCFESRNSLSLSIDSEDLLECCISSAMPKKKPSRLKGDNEK 2062
Db	1115	GTLTISIPSPSQSTIGSSQSTSPVSTTSGDMTSQG--STQIPGSTGTVTPQSTGS 1172
QY	2063	HSPRNNGGTLGLDLTLDLKIDIRPDSEHGLSPDSENFOWKAIQECa-----NSIVSS 2114
Db	1173	GSTISGEITSQGST-----QTPRSSLSTSPALSTSTQOQSVSTNSPGSTVTPQSTVGRS 1226

RESULT 8
t23327
adenomatous polyposis coli protein 1 - Caenorhabditis elegans
N:Alternate names: apr-1 protein
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23327; T03822
R:Gardner, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19727
A:Accession: T23327
A>Status: preliminary; translated from GE/EMBL/DBU

A:Molecule type: DNA
A:Residues: 1-1186 <N1>
A:Cross-references: EMBL:Z75712; PIDN:CA00045.1; GSPDB:GN00019; CESP:K04G2.8a
A:Experimental source: clone K04G2
R:Rocheleau, C.E.; Downs, W.D.; Lin, R.; Wittmann, C.; Bei, Y.; Cha, Y.H.; Ali, M.; Prie
Cell 90, 707-716, 1997
A:Title: Wnt signaling and an APC related gene specify endoderm in early C. elegans embr
A:Reference number: Z15051; MUID:97433081; PMID:9288750
A:Accession: T03822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1186 <R0C>
A:Cross-references: EMBL:AF013950; NID:G2338717; PIDN:AAC47747.1; PID:G2338718
A:Experimental source: Strain Bristol
C:Genetics:
A:Gene: apr-1; CESP:K04G2.8a
A:Map position: 1
A:Introns: 61/3; 113/2; 355/3; 551/3; 1067/3; 1161/3
A:Note: apr-1

Query Match 3.5%; Score 516.5; DB 2; Length 1186;
Best Local Similarity 21.9%; Pred. No. 9.1e-14;
Matches 289; Conservative 216; Mismatches 562; Indels 251; Gaps 52;

QY 271 MATSGNGCGSTTRMDHETASVLSSTHSAAPRLTSHLGTGYEMVYSLMLGTHDKDM 330
DB 1 MSSSSDENETT--IHRGTSGTGGGIYQPRAGSSKRTSNVRHVS-----DVIDE 50

QY 331 SRTLLAMSSODSCISMRQGCPLLIQLLHGNDDKSVLLGNRSGSKARASAAALHN- 389
DB 51 EEHVARF--REDTAEV--DDATVLLSLHFEKRDIVPTDEDDNK-----LRELHEK 100

QY 390 ---IHSQPDKKRREIRVHLLEQIRAYCETCWEQEAHEFGMDQKPNPAPVEHQI 446
DB 101 IFALITSESQVNRKRLKALPASNCVR-----EQVY--LRKPKSTPPASYHEL 149

QY 447 CPACVLMKLSDFEHRHANNELGGLOAIELLOVDCMYGL--TNDHYSITLRRYAGMA 504
DB 150 NAALHTIKESFGEEYRKAVTVLGLVEALAEVILEVHTFGINETPGEHRNIRKLIANA 209

QY 505 LTNLTFDGVANKATLCMKCMGRALVAQLKSESEDLQOVIASVLRNLNRADYNSKKTLLR 564
DB 210 LTNLTYQIHSKRRLCSYDGFRCVW--RVIESFNITQVYAGLIRNLNWSADSGMSEALQ 268

QY 565 EVGSVKALMECALEVKES--TLKSVLSALWNLNHAHTENKADICAVDGAFLAVLGTLYR 623
DB 269 P--TVHALSTAFAVHAHTRFDVTATLSALWNLNLAGHVENKRTICDTPNCLKVLASLLSP 326

QY 624 SQNTLAIISGGIIRNVSLIATNEDHQILRENNCLQTLHLKSHSLTIVSNACGT 683
DB 327 ARTSL--VDSATGILKYVQYLANSTHLEL--RSLLITRMLTLKSASFCTVNTLGA 382

QY 684 LNLNARNPKDQAL--WDGAVSNLKNL IHSKHMIAAGSAAALRNLMANRPAYKQANI 742
DB 383 IANLIVKDPHMQMIRODMAVQQLANLNRNSRDDIRTAVKSVLNTL--NQPCSHRYGDM 440

QY 743 M--SPGSSLSLHVKKALAEADAHLSETFNIDNLSPKASHRSKQRHKSGLYGDYF 801
DB 441 SHVGGGATGQMLSEPOLQWQ--TSHAYHGTASPRLLSLRATSPGKIYQQAQQQLI 499

QY 802 DTRNHDNRSD---NFTGMNTVLSPLYNTVTLPPSSSSRSGSLDSRSBK--DRSLEREG 857
DB 500 QTPQVDQRSSSLPRHF-----AVORNGFVMAQSVNQ--MDQHQQQQMIYQLQQQQQ 549

QY 858 IGLGNVHPATENCTSKRGLQISTTAAQIAKYMEEVSAIHTSOEDRSSGSTTELHCV-- 915
DB 550 I-----MTEDQAQMEHH-----QQIMYLQOQQQQFHQIQOQQQMKQAQADPVP 594

QY 916 TDERNALRRSSAAHTSNTYFTKSENSNRCTSMYPKLEYKRSSNDLSNYSNDGYOK 975
DB 595 TDDLDIPTVMTGRNSRSLGSMNPGSVMTNMNSSLDTAANSRSLSPSYND--IPA 653

QY 976 RGQMKRSIESYEDDSKFKCSYQYFADLAHKH-----SANHMDNDGELDPIN----- 1026

DB 654 SPTMCAQVFNLPKSTESE---HHQLTSQOQNTTHYSSGSAANTMTSRDGAITVPMDNIITP 710
QY 1027 -YSLK---YSDEQNSGRQSPQSNERWAPKHIITDEIKQSEQ-----RQSRNQS 1072
DB 711 TVAILNPLVHEQTPNGTVPKTSSELDSPDVLPGPSLEBEGDYAIIGAAOKTODEL 770
QY 1073 TTPVYTE---STDDKHLKFPHEG-----QBCVSPYRSRGANGSETNRV--GSNHGI 1121
DB 771 LTRSIQSEMPSTSTPKKVPRLNGFPSTQKTTSPPAWSH-----PDTSPIPKSSSHRT 826
QY 1122 NONVSQSLQDDY-----EDDKPTNYSERYSEEBEHEEPTNYSYKYNBEKHHVD 1174
DB 827 QPNRRQDASDADRLIMESIMSEMPKRIISPLAGTQCYLEPE--PERRSHSKNBEADRRD 885
QY 1175 -----QPIDYS-LKYATDIPSSQKQSFSSKSSGSSKTEHMSSESSENTSTPSSNAKR 1227
DB 886 AFTASHEPDSHNGIDVARGSDWSPQQQLHMESELESQASSEDSFGLTAEFPNSSTGA-- 943
QY 1228 QNLHPSSAQRSGQPOKAATCKVSSINQETIQTQYVEDTPICFE----- 1272
DB 944 -----AANTMRFDDEIDASLPWDCVDDDDYDTHDFEDYEDEEDPDPA 986
QY 1273 -----RCSLSLSLSAEDEIGCNOTTOEADSANTLQIAELKGIKIGTRSAEDP 1319
DB 987 TQFDGVDQAQTIDCSMTSGSSQR---NETTTTSDRSKALATSTPKGS----- 1034
QY 1320 VSEVPAYSQHPRTKSRRLQSSLSSESARHKAVFFPSGA---KSPSK-SGAQTPKSP-- 1373
DB 1035 ASSLPGVQATR-----VSTNGKSRLPVZKTNGSLVDKPKPIIASRRPLPKP 1084
QY 1374 -----EHVQTPLMFRCTSVSLDSFERS-----IASSVQSPGSGMVSGIISPS 1421
DB 1085 TLLKDKHYPEE-----DSIENQTRDDTIYNAPVVEAEQRIYMKALQOK 1130
QY 1422 DLPDSP--GQTMPSPRSKTPPP-----PPOTAQPKREVPKNKAPTAEKRESGPKQAAV 1472
DB 1131 NIEQSPSICNGSPIAKSAIVTPYNYQKPTPTGRNNGEMSNEKSVT-----PNPKQMLV 1183

RESULT 9
T23330
hypoetical protein K04G2.8b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23330
R:Gardner, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: 219727
A:Accession: T23330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1188 <N1>
A:Cross-references: EMBL:Z75712; PIDN:CA00048.1; GSPDB:GN00019; CESP:K04G2.8b
A:Experimental source: clone K04G2
C:Genetics:
A:Gene: CESP:K04G2.8b
A:Map position: 1
A:Introns: 61/3; 113/2; 355/3; 551/3; 1069/3; 1163/3

Query Match 3.5%; Score 514.5; DB 2; Length 1188;
Best Local Similarity 21.9%; Pred. No. 1.1e-13;
Matches 289; Conservative 216; Mismatches 564; Indels 249; Gaps 52;

QY 271 MATSGNGSTTRMDHETASVLSSTHSAAPRLTSHLGTGYEMVYSLMLGTHDKDM 330
DB 1 MSSSSDENETT--IHRGTSGTGGGIYQPRAGSSKRTSNVRHVS-----DVIDE 50

QY 331 SRTLLAMSSQDSCISMRQGCPLLIQLLHGNDDKSVLLGNRSGSKARASAAALHN- 389
DB 51 EEHVARF--REDTAEV--DDATVLLSLHFEKRDIVPTDEDDNK-----LRELHEK 100

QY 390 ---IHSQPDKKRREIRVHLLEQIRAYCETCWEQEAHEFGMDQKPNPAPVEHQI 446
DB 101 IFALITSESQVNRKRLKALPASNCVR-----EQVY--LRKPKSTPPASYHEL 149

QY 447 CPACVLMKLSDFEHRHANNELGGLOAIELLOVDCMYGL--TNDHYSITLRRYAGMA 504
DB 150 NAALHTIKESFGEEYRKAVTVLGLVEALAEVILEVHTFGINETPGEHRNIRKLIANA 209

QY 505 LTNLTFDGVANKATLCMKCMGRALVAQLKSESEDLQOVIASVLRNLNRADYNSKKTLLR 564
DB 210 LTNLTYQIHSKRRLCSYDGFRCVW--RVIESFNITQVYAGLIRNLNWSADSGMSEALQ 268

QY 565 EVGSVKALMECALEVKES--TLKSVLSALWNLNHAHTENKADICAVDGAFLAVLGTLYR 623
DB 269 P--TVHALSTAFAVHAHTRFDVTATLSALWNLNLAGHVENKRTICDTPNCLKVLASLLSP 326

QY 624 SQNTLAIISGGIIRNVSLIATNEDHQILRENNCLQTLHLKSHSLTIVSNACGT 683
DB 327 ARTSL--VDSATGILKYVQYLANSTHLEL--RSLLITRMLTLKSASFCTVNTLGA 382

QY 684 LNLNARNPKDQAL--WDGAVSNLKNL IHSKHMIAAGSAAALRNLMANRPAYKQANI 742
DB 383 IANLIVKDPHMQMIRODMAVQQLANLNRNSRDDIRTAVKSVLNTL--NQPCSHRYGDM 440

QY 743 M--SPGSSLSLHVKKALAEADAHLSETFNIDNLSPKASHRSKQRHKSGLYGDYF 801
DB 441 SHVGGGATGQMLSEPOLQWQ--TSHAYHGTASPRLLSLRATSPGKIYQQAQQQLI 499

QY 802 DTRNHDNRSD---NFTGMNTVLSPLYNTVTLPPSSSSRSGSLDSRSBK--DRSLEREG 857
DB 500 QTPQVDQRSSSLPRHF-----AVORNGFVMAQSVNQ--MDQHQQQQMIYQLQQQQQ 549

QY 858 IGLGNVHPATENCTSKRGLQISTTAAQIAKYMEEVSAIHTSOEDRSSGSTTELHCV-- 915
DB 550 I-----MTEDQAQMEHH-----QQIMYLQOQQQQFHQIQOQQQMKQAQADPVP 594

QY 916 TDERNALRRSSAAHTSNTYFTKSENSNRCTSMYPKLEYKRSSNDLSNYSNDGYOK 975
DB 595 TDDLDIPTVMTGRNSRSLGSMNPGSVMTNMNSSLDTAANSRSLSPSYND--IPA 653

QY 976 RGQMKRSIESYEDDSKFKCSYQYFADLAHKH-----SANHMDNDGELDPIN----- 1026

```

Db 101 IFALITSESDVNRKRLKALPASNCVR-----EQVY--LRKPSPTPASYHRL 149
QY 447 CPVAVLMLKSFDEHHRHANNELGGLQAIJAEILQVDCMYGL--TNDHYISITLRRYAGWA 504
Db 150 NAALHTIVKSGFGEYKRVATVGLVLAALAEVLLEVHTFGINETPGEHNRKLIANA 209
QY 505 LTNLTFGAVANKATLCSMKGMALVAQLKSESEDLQOVIAVLNLNSWADVNSKKTLR 564
Db 210 LTNLTYGQIHSKRRLCSYVDGPIRCV--RVIESPNITQVYAGLIRNLNSWADSGMSALQ 268
QY 565 EVGSKALMECALEVKES--TLKSVLNLNLNSHCHCTENKADICAVDGAFAFLVGLTYLR 623
Db 269 P--TVHALSIAVAHAHTRFDVATLSALMNLGHSVENKRTICDTPNCLKVLASLSPD 326
QY 624 SQTNTLAIIEGGGILRVNSSLIATNEDHRIILRENNCLQTLLOHLKSHSLTIVSNACGT 693
Db 327 ARFTSL--VDSATGILKYVSOYLANSTHLEL--RSLLIIRMLTLLKSAFTCVNTLGA 382
QY 684 LWNLSARNPKDQOAL--WDMGAVSMLKMLIHSKHMIANGSAAALNLMANRPKAKDANI 742
Db 383 IANLIVKDPHMQOIRQDMAVOQLNLVLRNSNRDITRAVKSVLNTIL--NQPCSHRYGDM 440
QY 743 M--SPGSSLPVLRKOKALBAELDAQLSETFDNIDNLSPKASHRSKQRHKQSLYGYDF 801
Db 441 SHVGGGATGMQLSEPOLQM--TSHAYHGTASPRLLSLRATRASPGKYIQPOAQOOLI 499
QY 802 DTRHDDNRSP---NFNTGNVTLSPLYNTVLPSSSSRGLSDSRSEK--DRSLERBERG 857
Db 500 QTPQVDORSSSLPRHF-----AVORNGFVMAQSYNOQ--MPQHQQOQMIYLOQQOQ 549
QY 858 IGLGNHYHATENPGTSSKRGQISTTAAIAKWEVSAIHTSOEDRSSGSTTELHCV-- 915
Db 550 IMF-----QTEQQAQMEH-----QOIVYLOQQOQQOQFIQOQQOQMAQADPVPP 596
QY 916 TDERNALRRSAAHNTNTFTKSENRTCSMPYAKLEYKRSNDSNLSNVSSNDGVGK 975
Db 597 TDDDLDTFTVMGTRNSERSLGMNPGSVMTNWNSSLDTAANSRALSPVSYND--IPA 655
QY 976 RGOMKPSIESVEDDESFCYQYPADLAHKLH-----SANHMDNDGELDTPIN----- 1026
Db 656 SPTMCAQVFNLPKSTES--HHQUTSQOQNTHYVSSGSANTMRSDGATTVPMDNIITP 712
QY 1027 -YSLK--YSDBELNGSRQSPQNRWARPKHIIEDETKQSEK-----PQSRNQS 1072
Db 713 TVAILNPLVHEQTNGTVPRKTSBELSDPDVLPGLSEEEGDIYALIGAAKTDDEL 772
QY 1073 TTPYVTE---STDKKHLKQPHFG-----QOECVSPYRGRGANGSETNRV--GSNHI 1121
Db 773 LTRSIQSEMPRSSSTPKVKVSPRLNGFFSPTOKTTSPPAWSH-----PDTSPKSSSHT 828
QY 1122 NQNVQSLSLQEDDY-----EDDKPTNYSERYSEEOHEBERPTNYSIKYNEEKHVD 1174
Db 829 QNRRQDASDADRLMESIMSEMPKRIISPLAGTQVLEPE--PERSHKNSEADRD 887
QY 1175 -----QPIDYS--LKYATDIPSSQKQSFSSKSSGQSKTEHMSSESSENTSTPSSNAKR 1227
Db 888 AFTASHEPDSHNGIDVARGSDWSPQQLHRMESLESQASSEDSFGLTAEPNSSTGA-- 945
QY 1228 QNQLHPSSAQSSGQPKAATCKVSSINGETITQTCVETPCFS----- 1272
Db 946 -----AANTVRFDEIDASLPMDCVDDDDYDYTDHFEDYEEDPDPA 988
QY 1273 -----RCSLSLSLSAEDEIGCNQTTQEAUSANTLOIAEIKGIGITRSABDP 1319
Db 989 TQFDDGVDQAQLIDCSMISGSGSSOR---NETTISRDSKALATSTPKGS----- 1036
QY 1320 VSEVPVAVSHPTKSRSLGSSLSSESARHKAVEFPFGA---KSPSK--SGAOTPKSP-- 1373
Db 1037 ASSLPVGRQATR-----VSTNGKGRLPVPKTNGSLVDKNPKPIIASRRPLPPK 1086
QY 1374 -----ZHYVQETPLMFSRCTSVSSLDSPSRS-----IASSVOSEPCSGMVSGIISPS 1421

```

```

Db 1087 TLLKKKHPEE-----DSIENQTRDDTIYVNAFVVEAEQERIYMNALKQOK 1132
QY 1422 DLPDSP--GQTMPPRSKTPPP-----PPQTAQTKREVPKKAPTAEKRESGPKAAV 1472
Db 1133 NTEQSPISGSGPIAKSAIVTPYNYQKPPFTGRNNGEMSEKSVT-----PNPKQMLV 1185

RESULT 10
A43359
microtubule-associated protein MAP1A - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C:Accession: A43359; S22108
R:Langkopf, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
J. Biol. Chem. 267, 16561-16566, 1992
A:Title: Microtubule-associated proteins 1A and 1C2. Two proteins encoded in one message
A:Reference number: A43359; MUID:92355629; PMID:1379599
A:Accession: A43359
A:Molecule type: mRNA
A:Residues: 1-2774 <LAN>
A:Cross-references: GB:M83196; NID:G205537; PIDN:AAB48069.1; PID:G205538
A:Note: sequence extracted from NCBI backbone (NCBIN:111039, NCBIPI:111040)
R:Cravchik, A.
submitted to the EMBL Data Library, June 1992
A:Reference number: S22108
A:Accession: S22108
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 73-364, 'NRURS', 370, 'QKN', 374, 'PSPKGL', 381-751, 'RSMMSOMNAOR', 764, 'D', 766, 'LRI', 'WLRNMCPPROSP', 851, 'V', 853, 'NSL', 855, 'LPHRWLKTN', 865, 'W', 867, 'HSQDPGGD', 877, 'Q', 879, 'I'
A:Cross-references: EMBL:X66840
A:Experimental source: strain Sprague Dawley
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein

Query Match 3.3%; Score 477; DB 2; Length 2774;
Best Local Similarity 18.7%; Pred. No. 1.2e-11;
Matches 542; Conservative 359; Mismatches 1039; Indels 958; Gaps 133;

QY 609 VDGALAVLTLYRSOT---NTLAIIE--SGGGIL-----RVSSLIATN 649
Db 1 MDGVAEF-----SEVSEITVDVPSFPDLLEPPISSGFLKLSKPCYIFPGGRGSAFVAVN 56
QY 650 -----EDHROILRENNCLQTLLOHLKSHSLTIVSNACGTLWNLSARNPKDQALWD 700
Db 57 GFNILDVGGSDRK-----SCFWKLVRHL-----DRIDSVLLTH 89
QY 701 MGAVSM--LKNLIHSKHMIANGSAAALNLMANRPKAKD--ANIMSPGSSLPVSLHV-- 754
Db 90 IGADNLPINGNLLQK-----VALEEEQSGSSISDWTNKLISPLGVVFNVPD 141
QY 755 -----RKQKALEAEALDAQHLSETFDNIDNLSFKASHRSKQR---HKQSLYGYDF 801
Db 142 KLRLPDASRAKRSIEACLTQLHNLRLGIQAEPLRVVSVNTIEPLTLFHKMGVGRLDY 201
QY 802 DTRHDDNRSDNF-----NTGNMTVLSPLYNTTV-----LPSSSSSR 838
Db 202 VLPVKDSKEMQMLQMKWAGNSKAKTGIVLANGKEAIEISVPYLTSTALVWLPANPTEK 261
QY 839 -----GSLDSSR---SEKDRSLE-----RERIGLGLG---NYHPATENPGTSSKR 876
Db 262 IVRVLPFGNAPQNKLEGLKLRHLDFRYPVATQKDLAAGAVPANLKPSKIHRADSKE 321
QY 877 GLQTS--TTAAQIAK---VMEVSAIHTSOEDRSSGSTTELHCVTDERNALRSSAAHTH 931
Db 322 SLKAAPKTAVSKLAREEVLKEGAKEARSELAKTEKKAPEKPEKPEKSKSERV 381
QY 932 SNTYNFTKSEN-----SNRTCSMPYAKLEYKRS--NDSLNSVSNDCY----- 973
Db 382 GESSEALKAEKRLIKDKACKKHLKEKLSLEKKEKKEKKEKELKEKELKEKKEK 441
QY 974 -----GKRGQMKPSIESYSEDDSKFCYQYPADLAHKLHSAHMDNDGELDTPINYS 1028

```

Db 442 DAKDEKRDTPKVKXLSKPDLPKF-----TPEVRKTLXKAKAPGRVKVDKGRA 491
QY 1029 LKYSDQNLNSGRQSPSQNE-----RWARKPHIED--ELKQSEQRQSNQST 1073
Db 492 ARGEKUSSEPTPAQKAAPAAVSGHRELALSSPDLTQDFEELKREERGLLAEQRD 551
QY 1074 T-----YPVYTESTDDKHLK-----FQF-----HFGQECVSPYRSRGANGSETNR-- 1114
Db 552 TGLGEKPLPADATQCHPSAAIQTQSGPVLGEHVEREKEVVP--DSPDKKS--TNRGP 609
QY 1115 -----VGSNHGINONVSQLCQBEDDYEDDKPTNYNSRYSEBEQ 1152
Db 610 DSGAEVEKEKETWBERKQREAEGLPENTAAARESEAEVKEVDIEKAELEEMETHPSDEE 669
QY 1153 HEEBERPTNYSIKYNEEKRHVDOPIDYSLKAYATI-----PSSOKOSFSSKSSG 1203
Db 670 GEETKAESFY-----QKHTQEAALKASPKSREALGGRDLGFGOKAPEKETASFLSSLAT 722
QY 1204 QSSKTEHMSSENSTPSSNAKR--ONQLH-----PSSAOSRSQG-----PQ 1244
Db 723 PAGATEHVSYIQDETIPGYSETEQISDEBIHDEPDERPAPPFRPFTSTYDLSGPEGPGPF 782
QY 1245 KAATCKVSSINQETIQYCYVEDTPICFSRCSLSLSSAADEIGCQNTTQEADSAANTLOI 1304
Db 783 EASQAADSAPASSSKTYGAPETELTY--PNVVAAPLAE--HVSSATSITE 832
QY 1305 AEIKGKIGTRSAED--PVSEPVAVSOHPRTKSSRLQSSLSSESARHKAVE--FPS-GA 1358
Db 833 CDKLSSPATSVAEDQSVASLTAQTEETGKSLLDVTWISPSRTEATQGLDYVPSACT 892
QY 1359 KSPSKSAQTP--KSPPEHVQVETPLMFSRCTSVSSLDSP-----SRSIASS----- 1404
Db 893 ISPTSSLEEDKGFKSP-----CEDFSVTGESEKKGTVGRGLGSEKAVGKE 939
QY 1405 ---VOSEPCSGMVSGIISPSDLPDPGOTMPPS-----RSKTPPP 1441
Db 940 EKVVTSEKLSGOYAAVFG-----APGHTLPGEALGEVEBERCLSPDSTVKAASPPP 993
QY 1442 --PPQTAQTVREVPKNAKPAEKR-----RVQVLPDADTLHFAFESTPDGFCSSSLA 1508
Db 994 SGPPSAHT--PFHQSPVEDKSEPRDFQEDSWGETHKSPGVSKEDSEETQVTKPGPEG 1049
QY 1471 -----AVNAAVQ-----RVQVLPDADTLHFAFESTPDGFCSSSLA 1508
Db 1050 TSEEGKPPTRSOAQDMPVSIAGGTGCTIQLPEQDAIVFETG-----EAGSNLGA 1103
QY 1509 LSL-----DBFFIQKVELR-----IMPPVQENDNGNETESEQP 1542
Db 1104 GTLPGEVRTSTERATEF--QKDEVLFTQSLSPEDAELSLSVLVSPDTTQEAATPRSP 1161
QY 1543 KESNENQKEAEKTIIDSEKDL-----LDSDDDDIEILERCIISAMPTKSRKKGKPAQTAS 1599
Db 1162 CSLKEQO-----PHKOLWMPSPEDTQSLSFSEZ-----SPSKET-----SLDISS 1202
QY 1600 KLPPPPVA-----RKPSOLPVYKL-----LPSQNRLOPKHVSPTPGDD 1637
Db 1203 KQLSPSSLGLQFGLNLEKGEKGPVWKAEDDSCHLAPVSIPEPHRAIVTSPDTETPAGT 1262
QY 1638 MP-----RVYC-----VEGTPINFSTATSGLTIESPPNELAA-----GE 1673
Db 1263 LPQGSFSGHALSVDRKHSPEBITGPGHEMT-----SDSGLTKSPESLSPAMEDLAVWE 1318
QY 1674 GVRGAQSGSEFERDT-----IP-----TEGRSTDEAGGNTSSVTIPE 1712
Db 1319 GRAPGKEPELKETSETROQKQILPEKAVVRODILIIHQDGLALDEENKPGEQDXT--PE 1377
QY 1713 LDNKAEBGDIILABCINSAMPKGGKHKPRVKKIVDQVQOASASSAPNKGDLQKKXP 1772
Db 1378 QKGRDLDEKTAALDKGPEPK-----BKDLREDQGORAGPPAEKOKASQORDT 1428
QY 1773 TSPVPIQNTVTRVRKVDKSNKNLNAERVFSNKKQKQNLKNNKDFNDKLPNNED 1832
Db 1429 LQQT-----QATEPRDAQERRD-----SEKDKSLELDRPTPEKDRILVQED 1472

QY 1833 RVRGSAFDSPHHYTPIEGTYPYCFSRNDSLSLDFDDDDVLSREKAELRKAKE---NKE 1889
Db 1473 R-----APEHSIP-----EPTQTDRAFPDRKGTDDKEQKEASEBEKQVLEKQDWALGKE 1521
QY 1890 SEAKVTSHELTSNCOASANKTQAIKQPIKQKPIKQKOSTPQSSKDIPIRSGAA--- 1946
Db 1522 GET-----LDQEARIAEQKDELKEDKTQGGKSSFVEDKTT--TSKETVLDQKSAKA 1572
QY 1947 TDEKLQNFALIENTPVCFSHNSSLSLSDIDQENN--KENEPI---KETEPPDSQGBP-- 1999
Db 1573 DSVEQDGALEKTRALGLEBSPAEGSKARQEKYKWEQDVVQWRETS--TRGEPVG 1630
QY 2000 ---SKPQASGYAKS--FHVEDTVPVCFSRNSSLSLSDIDQEN--SSAMPK--K 2050
Db 1631 CQKEPVPAWEGKSPQEVRVYWRDRDITLQDAYWRELSCDRKVWFPHELDQGGARPYCE 1690
QY 2051 KKPRL--KGDNE-----KHSRNMGGIIGEDLTLDKDIQRDSEHGLSPQSENEDWK 2102
Db 1691 ERESTFLDEGDEQEIITPLOHTPS-----PWTSDFKQFQEPLOKGLEVER---WL 1739
QY 2103 AIOEGANSIVSLHQAACLAASCLSQASSDSLSLSKSGISLSPFHLTPDOBEK---P 2158
Db 1740 A-----ESPVLPEEEDKLTTRSP 1758
QY 2159 F-----TSNKGPRILKPEKS-----TLETKKIESEK----- 2186
Db 1759 FEILSPASPEMTGORVPS--AFQOESPVPDTESTAPMNEPTTPSWLAIEIPWPVKDR 1816
QY 2187 -----GIKGGKVKYKSLITGKVRNSNEISGQKQPLQANMP 2222
Db 1817 PLPPAPLSPAPAPPTPAPEPHTPVFWSGLAEYDVVAQVQEALEGGPYSPGLKDYR 1876
QY 2223 SISRGRMIHLPVGRNSSSTSPVSKKGPLKT-----PASKSPSE----- 2263
Db 1877 KAEGERECEGAGAPDSSSPKVPKPEAGESLATRDTQTEPEQREPTPYDERSFQADI 1936
QY 2264 -----QQTATT--SPRGAK-----PSYKSELSPVARQTSIGGSKAPSSGSRDST 2308
Db 1937 YEOMMLTGLGPACTREPPPLGASGDWPHLSTKEAAGCNTSAKETSSPASPNLQSDT 1996
QY 2309 P-----SPPAQOPLSRPIQSGRN--SISQ-----RNGISDPNKLSQLPRT 2348
Db 1997 PAFSYASLAGPVPPOEP--DPGNVPEPITPPVPPRAPISLSDKSLPPLNGSTVSCS 2054
QY 2349 SSPSTASTKSGSGKMSYTPSGRQMSQNLTKOTGLSKNASSIPRSESASKGLNQ---MN 2405
Db 2055 PDRTPSPKETGRHW-----DDGTNDSLEKGAQEPEKXETRSPPSHHPMPWG 2103
QY 2406 NNGANKKVELSRMSSTKSSGS----- 2427
Db 2104 HSLWPETEAYSLSSDSHLGVSRLDPFASAFGFSLLQAPAPPQLPSPAPERSAPCGSL 2163
QY 2428 --ESDRS-----ERFVLVROSTFIKEAPSPTLRKLEESASPFES---LSPSSRRASP- 2474
Db 2164 AFSCDRAIALVPQTPTFTRHDEYLEVTKAPSLDSSLPLQPSPPSGPGLLSNLPRASFA 2223
QY 2475 ---TRSOATPVLSPEL-----PDMSLSTHSSVQAGG-----WRKLPNLSPTIY 2517
Db 2224 LSEGSSEATPVISSVAERFPFPLEAQAQSBGLSGKESAAHSLWDLTP--LSPA--- 2278
QY 2518 NDGRPAKRHDIAHSHSPSLPINSWTGKREHSHKSSLP-----RVSTWR 2566
Db 2279 ---PSNLDLAPAPAPAPAP-----GLPGDLGDGTLPCRPECTGEL 2318
QY 2567 TGSSSSILSASSSEKAKSEDEKHVNS---ISGTKQSKENQVS---AKGTWRKIKENEF 2620
Db 2319 TKKPSPLSPSGD--HEANGPGETS LNPPGFVTATAKEEAEPHAWERGSWPEGAERSS 2376
QY 2621 SPIN--STQSTVSSGATNGAESKTLIYQ-----APAVSKT 2654
Db 2377 RPDULLSSEQLPBGKSSGPPCSSLSEVEAGPQCATDPRPHCGELSPSFLNPLPPT 2436

QY 2655 EDVWRIEDCI-----NNPRSGR-SPTGNTTTPVIDSVSEKANPNIKSKDNQAKQNVGNG 2709
Db 2437 DSDLSLSTEEARLAGGRRRVRGPATGGPCWAD-----ETPPTSASDSSGSDSDVPPE 2493
QY 2710 SVPMRTVGLNRLTS-----FIQVD-APDQKGT-EIKPGQN-NPVPVSETNESP-----1V 2757
Db 2494 TEECPSTIAEAAALSDSDGDFLPVDKAGVSGTHHPRGHDPPPTPLDPRPPRPDVC 2553
QY 2758 ERTFSSSSSS-----KHSSPGTVAARVTPN-----YNPSAKSADSTS- 2799
Db 2554 MADPEGLSSSEGRVERLEKORPGRAPRAKPAAPARRLDIRGKSGTTPGKGVDRTSR 2613
QY 2800 --ARPSQIPTPVNNNTK 2815
Db 2614 TVPRSTPSQVTSAEK 2631

RESULT 11
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favellio, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match 3.1%; Score 452; DB 2; Length 3507;
Best Local Similarity 19.7%; Pred. No. 1.7e-10;
Matches 458; Conservative 275; Mismatches 859; Indels 738; Gaps 96;
QY 1051 RPXHIIEBIKQSEQRQSNQST-----TYPVVTSTDDKHLKFPQHFGQECV 1099
Db 629 RGNHSTSDLIITQVVQQRNFSTGTIILTRGVSSGEAVTQTTDAEF----- 676
QY 1100 SPYRSGANGSETRVGNHNGINQVSQLCEDDDYEDDKPTNYSEYSEEPCHEE----- 1155
Db 677 -----GLEISAADLAGSGGTLPTT-----LEPKIEGSKKASGGVWTEDEGEDEDIM 726
QY 1156 EERPTNYSIKYN-----BEKRVDOO----- 1176
Db 727 EEGSGMSITTINGTGITGSPRSEGTIRVITTLGEDGEPETATKPGISAPDKTGEKSKE 786
QY 1177 IDYSLKYATDIPSSQKQSFSSKSSGQSKTEHMSSESTSTPSSNAKRONQLHPSSA 1236
Db 787 SDGEEKLITVEKDGKEAQSSGSSATSSGKKSEATSGSSSSSAKEGTGSEAGSGGASSSSG 846
QY 1237 QSRSGQPKA-----ATCKVSSINQETIQYCYVEDTPICFSRSSLSLSSAEDEI- 1287
Db 847 SGVSGSGSVSTESGSGFTSSGSGVSGEATGTGTGDSGSE-----SGKPSKSTEEKLP 900
QY 1288 -----GQNQTQEDASNT-----LQI 1304
Db 901 FTXNGEKSPISGSDTTTGKESSEBETTRKPIEGSDSLTEGSGGEWFTGSKGHFEGSKV 960
QY 1305 ABIKGIKGTAEADPVSEVPAVSHPRKTSRRLQSSLSSESARHKAVFP----- 1355
Db 961 SVTSGKPTQSGAEGSGGPKVPKPGCAPBITTDGESSSTSTGDKSGGPKADKSNKVV 1020
QY 1356 --SCAKSP-----SKSGAOTPK-----SPPEHYVQETPLMFRCTSVSLDSF 1396

Db 1021 PXTGDKNPDIITTDGEDSTSETSGGQGPKGKSGQPGDGKGEVKKPTSEVDGPQNLSGT 1080
QY 1397 ESRSTASSVQSEPCSGMVGSIISPSDLP-----SPG-----OTMPPSR 1436
Db 1081 KGKS-----NVP-----LKPIDLEEGSGIITSSGGKNSFEHCTKLERLPP--- 1123
QY 1437 KTTTTPPQTAQTKREYRKNKAPTAE---KREGPKQAAVNAAVQVRVLPDADTLHLFAT 1493
Db 1124 KTEDKSEETPOLGLEISACKPEPEDGTSKEVG-----LEILWESTTTPGSTITL 1171
QY 1494 ESTPDGFCSSLSLSLSLDEPIQ-----KDVLELRIMPPVQVNDNGNE-- 1536
Db 1172 DSDVGLAISGSDLTATKTKPHVEIGSGTGDBEITATTRDVSKTKKPRVEVDGDNGCE 1231
QY 1537 -----TESQPKESNENQKEAEKTIIDSEK-----LLDDSDDDDIIFILECIIIS 1581
Db 1232 TSGVDCKPTTAPTTPSSAESSTSRIPTTSEASPEGSGEAGVPESPDGSGESSTAPDG 1291
QY 1582 AMPTKSRKGGKPAQTAASKLPP-----PVAKPSQLPVYKLLPSQNLQPKHVSFTF 1634
Db 1292 VSPSTATAPEVPTTSASSTPDVAESGIPSTSKPTAREPLETTAPSTEVTSPEG--SGTE 1349
QY 1635 GDDMPRVYCEGT-PINFSTATSLDITIESPNELAAGEGVR-----GGAQSGE 1683
Db 1350 ESTLPP---TEGSGESTTSAPTVEPATVLPQNRNEKEPTKDTFTALPTTTTGAPOAND 1406
QY 1684 FEKRDITPTEGRSTDEA-----QGG--KTSSVTIPE----- 1712
Db 1407 SSVENTKCT---SSDEGLDLCERRTGVCRCEPFGEGAPPKKSCVDVDECATGDHNCHE 1463
QY 1713 -----LDNKAEB-----GDLAEICIN 1729
Db 1464 SARCONVGVGACFCPTGFRKADDDGCGQDIDECTEHNSTCCGANAKVKNKPGTVSCECN 1523
QY 1730 SAMPKGGKSHKPRVYKIMDQVQOASASSAPNKN-QLD-----GKXKXKTSPTV 1776
Db 1524 GFLGDGVQCVF-TTKKPCDSTQSKSHCSBNMSCEVDTVDGSGVECKECCGGYKKGKVC 1582
QY 1777 KPQNTYEYTRVRKNADSKNNLNAERVSDNKDKSKQNLKNS--KDFND-----KL 1827
Db 1583 EDINECVAEKAPCSLNA---NCVNMNGTFS---CSCKQGRGDFGFMCTDINECDERHPCH 1636
QY 1828 PNNE-DRVRGSAFADSPHY-----TPIEGT-----PYCFS--RND 1861
Db 1637 PHAECTNLESGFCECHSGFEGDGIKKTNPLERSCEDVEKFCGRVDHVSCLSVRYNGS. 1696
QY 1862 LSSL-----DFDDDDVDLSREKAEKLA-----KENKESEAK 1893
Db 1697 LSSVCECEPGFRFEKESNSCVDIDECESRNCDPASAVCVNTEGSRVRCBAEGYEGGG 1756
QY 1894 VTSHELTSNQOSANKTOAIKQPINR-----GQPKPILOKQSTTFPQSSKDI 1940
Db 1757 VCTDIDECDRGMAGCDSNMC---INRGSGCKCMAGYTGDCGATCIKIEEE-PKS--- 1808
QY 1941 PDGAATDE-----KLQNFALNTPVCFSS-----H--NSSLSL--SDIDQENN-- 1980
Db 1809 -DKTACTDWSRLCELEKQCTVDEEVPOCGACLPGHHPINGTCQSLQISGLCAQKND 1867
QY 1981 NKENEPKETEPPDQSGEPKQASGYAPKSFHVEDTPVCFSRNSSLSSLSIDSEDLLO 2040
Db 1868 NKHAECI--DIHPDSH-----FCSCPQGF-IGDGMICDDVDECNAGMCDDEKTC 1916
QY 2041 ECIS3-----AMPKXKPPSR--LKGDNEKHSPPNMG-----GILGE- 2074
Db 1917 NTIGSFNCVLEGGFKVDEKCKVDEKQPNREKIEIDENSSSSNSGQEKPTTKGIVSST 1976
QY 2075 DLTLDLKDIOQPDSEHGLSPSENFDMKAIQEGANSIVSSILHQAAAACLSRQASSSDS 2134
Db 1977 SATSESTTAEFHVTTTST 2034
QY 2135 ILSLKSGLSLGSPFHLTPDQ--BEKPTSNKGR--ILKPGKSTLTETKKIESKGI 2188
Db 2035 TVS-----STFESSSSSEAPLTSPTTTEVITSSSVKST--TPKERSSE-- 2078

[illegible]

```

RESULT 13
T20532
hypothetical protein F07A11.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20532, T2777
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20532
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2722 <WIL>
A:Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b
A:Experimental source: clone F07A11
R:Gajadaty, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27777
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2722 <W12>
A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CESP:F07A11.6b
A:Map position: 2
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1
Query Match 2.9%; Score 428.5; DB 2; Length 2722;
Best Local Similarity 18.6%; Pred. No. 1.1e-09;
Matches 460; Conservative 324; Mismatches 893; Indels 801; Gaps 107;
QY 811 SDNFTNGMTVLSPYLNTTVLPSSSSSSGLDSRSEKORSLERRERGIGLGNYPHATENP 870
357 SDRNL-----YFINRKENIEVARSSSPTSKSENDQ-----GSSSP----- 394
DB 871 GTSKRGLOISTTAAQIAKYAEVSAIHTSQEDR---SGSTTELCVTDERNALRRSSA 927
395 --SSSRDRQLHDPLOTRSSVEH---HTNQEDQENNASGDS-----SSDSDEEGSSS 443
QY 928 AHTSNFTNYFTKENSNRKTCMPYAKLEYKRSSNDGLNSVSSNDGYGKGQMKPSTESYS 987
DB 444 SNEDSQDNVDDEDDDDVV-----EEKR--HEEPEGKSSSPGNGHRDESNG--DKDH 493
QY 988 EDDSEKFCSCYQYPADLAHKIHSANMDDNDGEI-----DTPINYLKYSDEOLNSGRS 1042

```


QY 2758 ERTFPSSSSSKHSSPS---GTVAARVTFNVPNPSRKSSADTSARPSQIPTPVNNT 2813
Db 2428 KOIPTSEEDTDDSKADSMCAEAGAFRIL-----SRSTWGNSSGSPASGT----- 2474
QY 2814 KRDSTKDTSTESGTSQSP 2831
Db 2475 --TSPSTSSISGSPDSP 2490

RESULT 14
E88320
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88320
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: AY5000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/usc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2738 <STO>
A:Cross-references: GB:chr_II; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A11
C:Genetics:
A:Gene: F07A11.6
A:Map position: 2

Query Match 2.9%; Score 426.5; DB 2; Length 2738;
Best Local Similarity 18.5%; Pred. No. 1.4e-09;
Matches 461; Conservative 325; Mismatches 891; Indels 817; Gaps 108;

QY 811 SNFNTGNMTVLSPYINTVLPSSSSRGLSDSSRSEKDRSLERIGICLGNYPATENP 870
Db 357 SDRFNL-----YFINKKENIEVAARSSPTSSENDQ-----GSSSP----- 394
QY 871 GTSSRGQIQISTAAQIAQIAKMEVSAIHTSQDR---SSGSTTEHCVTDNRALRRSA 927
Db 395 --SSSRDRQNLHDPLQTRSSVEH-----HTNQDQENNASGDS-----SSDSDEEGSS 443
QY 928 AHTHTNTVFTKSENSRNTCSMPYAKLEYKRSNDSLSNSNDGCGYKRGOMKPSIESYS 987
Db 444 SNEDSEQNDVDEDEDDVVS-----BEKR--HEPECKSSSPGNGHDSNG--DKDH 493
QY 988 EDESKFCSGYGYPADLAHIAHSANHMDNDGEL-----DTPINYSKYSDQLNSGRQS 1042
Db 494 EDSSERF---SQPSTSSH--HETSHSPKDESEAYQSRSFPLNY-----QS 534
QY 1043 PSQNERWARPKHIIEDETKQS--EQCRSNQSTTYPVYTESTDD-----KHLKTPHFG 1095
Db 535 QSPGYEFLEK-----EIKQEFSPITSSASSDLEDMEMPNDPLTRMLERNHWRPFID- 588
QY 1096 QECVSPYRSGANGSETNRVGNHGINQNVQSL-----QOEDDYED-----DK 1139
Db 589 ---VSSF-----VNRIDIVELNLCARASYEKFTRPPFKCNNDVLISQIKIVTHE 636
QY 1140 PTNYERYSEERQHEERPTNYSKYNEKHVDQPIDYSLKATDIPSSQKQSFPSK 1199
Db 637 PRDY--YYENPCSELEVRIDW-----RKLSDTADLD-----DFRATDSKELGRDQ 681
QY 1200 SSSGQSKTEHMSSSSENTSTPSSNAKQNL-----HPSSAQRSGQPKAATCKVSS 1253
Db 682 PAGGTS-----GRPSLDESRTNLSFPDTHHPALQRS----- 716
QY 1254 INQETIQYCV-----EDTPICFSRCSLSLSAEDIEGNCQTQOADSANTLOI----- 1304
Db 717 -----HSLCIGFMTPTSPF-----PTSQPLLVNTTHLPQTSG 748
QY 1305 ABIKIGITRGAEDPVSVPAVSQHPRTKSRLOQSSLSSESARHKAVEFPSSGAK----- 1359
Db 749 PSTSGIITPRSSQPPPLMSPVSRNMSST---GRPASIQTLRHQSVNFPDPVSPPPPP 805

QY 1360 -----SPKSGAQT-----PKSPPEHVHVQETPLMFSR 1386
Db 806 IPPTHDEMAPRGTPPPRRSRSETMVLRLSPFPCTPIQNLTLMTPIVPPPHLIAAT----- 859
QY 1387 CTSVSSILDSFESRSIAGSVQSEPCSGMVSGIISPSDLDPGQOTMPPSRKTIPTPPPPQTA 1446
Db 860 -----STGTHSVSSSAHSTP-RHSISG-----TPVH-CEFSNKTSGPP--TP 898
QY 1447 QTKREVPNKAPTAKRRESQPKQAVNAVQVQLVPADATLLHFATESTPDGFSCESSL 1506
Db 899 KSRPEKVIQIRHDTISK--SGFSN-AINALQARSQMTSGDP--KKSAPSTPVVRDAGSGL 953
QY 1507 SALSLDEPFFQKQVELRIMPPVQE-----NNGNET----- 1537
Db 954 VAQIMSN---QPNLGLRLKLPRIEKKSSALQNIQHQPPhNSANSTPSTPSTHQAAMFKD 1010
QY 1538 ESQPKESNENQPKAE----- 1554
Db 1011 KEKERKKKEKERERAREREMKRKETKEERNKEMERAKLEDERQERKEKKERD 1070
QY 1555 -----KTIDSEK-----DLDDSDDDIEILEECIIISAMPTKSSRKGKKA 1595
Db 1071 ERKEKEKVRKKAKEKELKKKKHKKGSDSDSDNDELDDLVRKS--TKEMTQEEKDH 1128
QY 1596 QTASKLPPPVARFSPQLPVYKLLPSONRLOPKXHVSTPGDDMPRVYCVVEGTPIINFSTAT 1655
Db 1129 QLALL-----SKGIIENLKSRRSRDKRAHDSFE-----KMQQKS 1164
QY 1556 SLSDLTTIESPPNELAAGEVGRGAOSGE---FKRD-----TIPTEGRSTDE---AQGG 1703
Db 1165 QQRVLIESDDE--GGKDGDKGNSNGEESDSEKADLPPPPAPFSLSESADQRLKVLKER 1223
QY 1704 KTSVTTPELDNDKAEEDILAEICINSAMPKSHKPFVRVKIMDOVOOASASSAPKN 1763
Db 1224 EKGLTITSSDEHDNAGELHQRLTEDRENKQKSLTAYSSDEQGERKXVFRWRDD 1283
QY 1764 QLOKKKKKTPSVKPIPTONTEYTRVRKNADSKNNLNAERVFS-----NKDSK 1812
Db 1284 SEDAAAKHPGWSAKDQKQRKLEHRRSEDESKNAKDRFDIPHEDVSDEETEDGS 1343
QY 1813 KONLKNNSKDFNDKLPNNEDVRVG-----SFAFDSPHHTPIEGTPYCF 1856
Db 1344 RSRQSTSTISNVATAKERKSGKTPLRIVPPTGTPLLSPKILSPKHUSPTST--SS 1401
QY 1857 SRNDSJSLDFDDDDVLSREKAEALRAKENKESEAKVTSHTELTNSQNSANKTOAIKQ 1916
Db 1402 TKRSSIS---DHENLISPRQ-----NRTTSST--TATTSSKHEALSIEPK 1443
QY 1917 PINRGQPKILOKSTFPQSSKDIP-----DRGAATDEKLQNF-----AIEN 1958
Db 1444 PLS-----PPVTAKSSV---SSIDDPSTIRDEFFSNMSAADSPMTTGRPMVLTAKAMKAFNS 1496
QY 1959 TPV-----CF-----SHNSSLSSLSIDID-----QENNNKENETPKET 1990
Db 1497 TPKKYVSYSLLIDCYMLGNMAKNSSGQHDSSGSSSDSSSDSSSDSSDDEVPKQT 1556
QY 1991 EPDPS-----QGEPSK-----PQASGVAPKSFHVEDTPVCFPSRNSSSLSSIDSEDDL 2038
Db 1557 EPVTSIPWASDNGSPENVVETFSVQTPR---EPSPFTTSEQS-----ESEPEA 1606
QY 2039 LQSC-ISSAMPKKKPSRLKGDNEKHSPRNMGIIGEDLTLDLKDIOQRPDSEHGLSPDSE 2097
Db 1607 VPCPPEASVEPQNETQNTVEPVSEHEDSHEHG-----DSEAVESQOQP-LEH--QEEKE 1659
QY 2098 NFDWKAIQEGANSIVSSLHQAAAAACLSRQASSDSLSLSKSGISLGS--PFHLTPDQE 2155
Db 1660 ELENKILD-----VAAHEHEHQVQGEDSV--ESSIPAPSDPEPVPVTOAQE 1703
QY 2156 EKFPFT-----SNKGPRLLKPE----- 2172
Db 1704 KSAHTLISDQETQAVQSIFDEEEADEFFQYPDFGISTNEKEVSGKDPHNHKTETPLANG 1763

2173 QY -----KSTLETKKIESBSKGIKGGKQV-YKSLITCKVRNSSEISGQMK 2214
 1764 DB HTDLLFPSSSAHASEKQSTKSEDDHEEDSELVMEKEVPMQVIAQEVHVPSEPSMEE 1823
 2215 QY -----PLOANPSPISRGRTMTHIPVRNSSSSTSPVSKGKPLK 2255
 1824 DB EVKLETSFVPEKBEFKMEESFEQPTFDLISNNESQDTPGAVNNH-----LHENHDAVOT 1878
 2256 QY PASKSP-SEGQTATTSRGAQPSVKSELSVAVQTSQIGGSSKAPSRSGSRDSTPSRPAQ 2314
 1879 DB PIQLOPASQHVQAPSRPA-----VAPDSQONGPV-----LVSQSQSPS--- 1918
 2315 QY QPLSRPTQSPGRNISPRNGISPPNKSQLPRTSSPSTASTKSSGSGKQVTSRQMS 2374
 1919 DB PMSQOSQDMAQNLLSSKDINDLAALKHKNPEALA---QATRGDCSGIFQHLLHAQGN 1974
 2375 QY QONLTQTKGLSKWASSIPRESASKGLNOMN-----NGNGANKKVELSRMSSTKSSGSE 2428
 1975 DB GONWTPENMLQKAFFAQOQENANQWQAKMKQQTINKDRIKEQERVRYEENKRYE 2034
 2429 QY SDRSER-----PVLVRQSTFIKEAPSPILRRKLE-----ESAS--- 2461
 2035 DB EDREKQKKEERQELAAATAAATATQKAAEALKQKQEVPRHGFQHLVSMMTPEARSILY 2094
 2462 QY -----FESLS-----PS---SRPASPTRSQAQTPVLSPLDPKXSLTHS--- 2497
 2095 DB EQPFLSSYINRDSIGATNGVLHFTQSIQRPSS-TASTSSNPPKAPLOPSPASVNONITD 2153
 2498 QY -----SVQAGGWRKLPNLSPTIEYNDGRPAKHDIARSHSESPSRLPINRSQTKWR--- 2549
 2154 DB PAETEEIRVQRFYKPLKMSAE-----EAATMAVASSDPNPATSTVDLAAMLQOLA 2207
 2550 QY -----EHSKHSLSIPRVTWRTGSSSSI---LSASSSESEKAKSEDEKHVNSIGTKQSKENQV 2606
 2208 DB AQAAQAAQVPPVVTASTPPLNLETLTSLANLATGALNPLSLMALTSSLNQOSP 2267
 2607 QY SAKGTWRKIKENESPTNSTSTQVSSGATNGAGSKTLIYOMAPAVSKTEDVWVRIEDCPI 2666
 2268 DB VYQIARVLLTMNGQMLATHQISELLATWN-QOETLMALLAAR-----NGLFF 2315
 2667 QY NNPRSGSP---TCN---TPPVIDSVSEKANPNIKOSKNOAKONGVNGVPMRTVGLEN 2720
 2316 DB AMPQONQOPQMPAQGGAFAITPVLPHMSLKEN-----AKD---QLSVGGVSDRKXSCPLHA 2367
 2721 QY RLTSFIQVDAPDQGTIKG-----QNNPVPVSETNE----- 2753
 2368 DB MIGQGOQPPPPQPMQAVAPAPRSPSPRKSMPENLPPEWKEKNEMFRKEILRLDIL 2427
 2754 QY -----SPIVERTPFSSSSSKHSKSPS-----GTVAARVTPPNVNPSPRKSADS 2797
 2428 DB LEEELGADEEDQKDLAQIPTSEEDTODSKADSMGABGSAFRIL-----SRSSMTGN 2480
 2798 QY TSARPSQIPTPVNNNTKRSKTDSTSSSGTQSP 2831
 2481 DB NSGSPSASGT-----TSPSTSSSISSGPDSP 2506

RESULT 15
 T02345
 hypothetical protein KIAA0324 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
 C:Accession: T02345
 R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
 submitted to the EMBL Data Library, March 1998
 A:Description: Sequencing of human chromosome 16p13.3.
 A:Reference number: Z14664
 A:Accession: T02345
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1791 <RIC>
 A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650

C:Genetics:
 A:Map position: 16
 A:Introns: 1610/2; 1706/2
 A>Note: KIAA0324
 Query Match 2.9%; Score 419.5; DB 2; Length 1791;
 Best Local Similarity 15.68; Pred. No. 1.5e-09;
 Matches 374; Conservative 259; Mismatches 742; Indels 537; Gaps 75;
 1210 HSSSSSENTS-----TPSSNAKRNQLHPSSAQSRSQOPQAAATCKVSSI 1254
 DB 1 HSGSSSDTKVKPETPRQSHSGSISPVVKVKAQTPPGPSLSGSKSPCP----- 49
 1255 NOETIQYCVEDTIFCSRSSLSLSAEDEICGNTTQEAADSANTLQIAETKIGTR 1314
 DB 50 -QEKSDSLVQSCFGSLCA-----GVKSSTPFSGSYFVGSLSLQKGO--SQ 94
 1315 SAEDPVEVPVAVSQHPRTKSRLOGSSLSSESARHKAVEFPSPGAKSPSKGAOTPKSPPE 1374
 DB 95 TSPHRSD-----TSPVQRQSHSESPSLQSKS-----QTSPPKGRSRSSSPVT 138
 1375 HYVQETPLM-----FSRCTSVSSLDSPESRSIASVQSEPCSGMVSGIISPSLDLPDQPG 1429
 DB 139 ELASRSPIRQDRGFSAAPMLKSGMSPEQSRFQSDSSSYFTVDS-NSLLGQSRLETAEK 197
 1430 ---TWPPSRKTPPPBQT-----AQTKEVPKNAKPTAEKRESGPK 1468
 DB 198 EKMALPQEDATAPPRKOKFSPFPVQDRPESSLVFKDLRTPPERSGAGSPETKEQ 257
 1469 QAAVNAARVQVVL-----PDATLLHFAT---ESTPDGFCSSLSL-----ALSLD 1512
 DB 258 NSALPTSSODEELMEVVEKSEPAQILSHLSSELKEMSTSNFESSPEVEERPAVSLTD 317
 1513 EPFTQKDVLEIMP-----PVQENDNGN---ETESQEPKESNEQ 1549
 DB 318 OSQOASLEAVEVPSMASSWGPHFSPEHKELNSPLRENSFGSPLFRNSGGPLCTEWT 377
 1550 EKEAKTTDSEKDLDDDDDDIIEECIIISAMPTKSSRKKKPAQATASKLPPPVARKP 1609
 DB 378 GFSSEVKEDNGPFLNQLTDPSLDMKE---QSTRSS-----GHSSELSPDAVEKA 426
 1610 SOLPVYKLLPQNRLQPOKHVSFTPGDDMPRVYCVGTPINFTATS--LSDLTIESPPN 1667
 DB 427 G-----MSSNQSISSPVLDAVPTPRERS-----SSASPEMKDGLPRTPSR 469
 1668 ELAAGE--GVGGAGSGFEKRDITPTTEGRSTDBAQGGKTSVITPBLDDNKAEEGDLA 1725
 DB 470 RSRSGSPGLRDGS-----GTPSRHSLSGSS-----PGMKDIPRTPSRGRS 510
 1726 ECINSAMPKSGKHKPFRVKTMQVQASASSAPNKNQ---LOGKKKKPTSPVKPIQNT 1793
 DB 511 ECDSPPEKALPQTP-----RPRSRSPSPELNNKCLTPQERSGSE--SSVQKT 559
 1784 EYRTEV--RKNADSKNLAERVFSDNKKONKONKNSKDFNDKLPNNEDRVGSPAFD 1841
 DB 560 VARTPQORSRSGSQELDVPSASQPERSE-----SDSSPDSKAKTR----- 602
 1842 SPHYHTPIEGTFCFSRNDLSLDDDDVDLSREKAELEKAKENKESAKVTSHTLT 1901
 DB 603 -----TPLR-----QSRSGSGSPVD-----SKRSLSPRRSRSGSPSEVADKPRAA 643
 1902 SNOQKANTQAIKAPINRGOPKPILOKQSTFPOSSK---DIPRGAATDEKLNQFAEN 1958
 DB 644 PRAQSGSSSEPKAPAPRALP-----RRSRSGSSKGRGSPGSGSSTES-----SPEH 693
 1959 TPVCFSHNSLSLSDIDQENNNKENEP-IKETBPDSQGPSPQASGYAPKSFHVED- 2016
 DB 694 P2-----KSTARRGSRSPPEPKTKSRTPRRSRSSRSPPELTKARLSRRSRA 742
 2017 -----TPVCFSRNSLSLSDSDEDDLQBCISSAMPKPKKPKRLKDNKHSRPN 2067
 DB 743 SSSPETSRTPPHRRRSPSVSPPEAKSRSSRRRRSSSPRTKTTTSR-RGRSPSPKPRG 801

Search completed: August 25, 2004, 17:24:44
Job time : 98 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:21 ; Search time 32.5 Seconds

(without alignments)
4554.937 Million cell updates/sec

Title: US-09-442-489F-2

Perfect score: 14575
Sequence: 1 MAASVDQLLQVEALKVEN.....ESSGTQSKRHSGLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	14533	99.7	2843	1 APC_HUMAN	P25054 homo sapien
2	13168.5	90.3	2842	1 APC_RAT	P70478 rattus norv
3	13148	90.2	2845	1 APC_MOUSE	Q61315 mus musculu
4	477	3.3	2774	1 MAPA_RAT	P34926 rattus norv
5	474.5	3.3	5560	1 SPEN_DROME	Q8x883 drosophila
6	445	3.1	3664	1 MINT_HUMAN	Q86t58 homo sapien
7	434.5	3.0	3644	1 MINT_MOUSE	Q62504 mus musculu
8	417	2.9	2688	1 ZEP1_MOUSE	Q03172 mus musculu
9	400.5	2.7	2805	1 MAPA_HUMAN	P78559 homo sapien
10	400	2.7	5085	1 PCLO_RAT	Q9jks6 rattus norv
11	395	2.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
12	387.5	2.7	5147	1 PCLO_HUMAN	Q9Y6V0 homo sapien
13	385.5	2.6	3210	1 CENF_HUMAN	P49454 homo sapien
14	381	2.6	2459	1 MAPB_RAT	P15205 rattus norv
15	380.5	2.6	4377	1 ANK3_HUMAN	Q12955 homo sapien
16	379.5	2.6	4911	1 MLL3_HUMAN	Q8ne24 homo sapien
17	375.5	2.6	3562	1 PGCV_CHICK	Q90953 gallus gall
18	374	2.6	5038	1 PCLO_MOUSE	Q9qyx7 mus musculu
19	373	2.6	3256	1 K167_HUMAN	P46013 homo sapien
20	371.5	2.5	3969	1 HRX_HUMAN	Q03164 homo sapien
21	371	2.5	2468	1 MAPB_HUMAN	P46821 homo sapien
22	369.5	2.5	2464	1 MAPB_MOUSE	P14873 mus musculu
23	368.5	2.5	2453	1 NCR1_MOUSE	Q60974 mus musculu
24	368	2.5	2440	1 NCR1_HUMAN	O75376 homo sapien
25	366	2.5	2492	1 ATRX_FANTR	Q7yqm4 pan troglod
26	365	2.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
27	365	2.5	3866	1 HRX_MOUSE	P55200 mus musculu
28	365	2.5	5120	1 PCLO_CHICK	Q9pu36 gallus gall
29	364.5	2.5	1581	1 PRFB_HUMAN	Q15648 h peroxisom
30	362	2.5	3259	1 GBL1_HUMAN	Q14789 homo sapien
31	360.5	2.5	6632	1 UN89_CAEEL	O01761 caenorhabdi
32	357	2.4	2492	1 ATRX_PONPY	Q7yqm3 pongo pygma
33	356.5	2.4	2663	1 CENE_HUMAN	Q02224 homo sapien

ALIGNMENTS

RESULT 1

ID	APC_HUMAN	STANDARD;	PRT;	2843 AA.
AC	P25054; Q15162; Q15163;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Adenomatous polyposis coli protein (APC protein).			
GN	APC OR DP2.5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
FP	SEQUENCE FROM N.A.			
RX	MEDLINE=91335210; PubMed=1651562;			
RA	Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,			
RA	Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,			
RA	Finleary R., Markham A., Groffen J., Boguski M.S., Altshul S.F.,			
RA	Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,			
RT	"Identification of FAP locus genes from chromosome 5q21.";			
RL	Science 253:661-665(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=91330307; PubMed=1678319;			
RA	Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,			
RA	Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,			
RA	Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,			
RA	Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,			
RA	Abderahim H., Cohen D., Leppert M., White R.,			
RT	"Identification of deletion mutations and three new genes at the			
RL	familial polyposis locus.";			
RL	Cell 66:601-613(1991).			
RN	[3]			
RP	ASSOCIATION WITH CATENINS.			
RX	MEDLINE=94082295; PubMed=8259519;			
RA	Su L.-K., Vogelstein B., Kinzler K.W.;			
RT	"Association of the APC tumor suppressor protein with catenins.";			
RL	Science 262:1734-1737(1993).			
RN	[4]			
RP	DISEASE.			
RX	MEDLINE=95174843; PubMed=7661930;			
RA	Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,			
RA	Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,			
RA	Wood P.A., Taqi F., Bocker S.V., Petersen G.M., Offerhaus G.J.A.,			
RA	Terstmette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.;			
RT	"The molecular basis of Turcot's syndrome.";			
RL	New Engl. J. Med. 332:839-847(1995).			
RN	[5]			
RP	DISEASE.			
RX	MEDLINE=97094176; PubMed=8940264;			
RA	Eccles D.M., van der Luijt R.B., Breukel C., Bullman H., Bunyan D.,			
RA	Fisher A., Barber J., du Boulay C., Primrose J., Fodde R.;			
RT	"Hereditary desmoid disease due to a frameshift mutation at codon 1924			
RT	of the APC gene.";			

Q62059 mus musculu
P46100 homo sapien
Q912u6 mus musculu
P08640 saccharomyc
P30415 mus musculu
P30414 homo sapien
Q12888 homo sapien
Q9wu42 mus musculu
P47035 saccharomyc
Q9nm4 caenorhabdi
O35923 rattus norv
P48415 saccharomyc

34 356 2.4 3358 1 PGCV_MOUSE
35 355 2.4 2492 1 ATRX_HUMAN
36 352 2.4 7389 1 BPAL_MOUSE
37 350.5 2.4 1367 1 AMYH_YEAST
38 350 2.4 1453 1 NKCR_MOUSE
39 350 2.4 1462 1 NKCR_HUMAN
40 350 2.4 1972 1 P531_HUMAN
41 349 2.4 2472 1 NCR2_MOUSE
42 345.5 2.4 1189 1 YUHE_YEAST
43 345.5 2.4 8545 1 ANCI_CAEEL
44 345 2.4 3343 1 BRC2_RAT
45 344.5 2.4 2195 1 SC16_YEAST

RA Am. J. Hum. Genet. 59:1193-1201(1996).
 RN [6]
 RN DISEASE.
 RP MEDLINE=20243021; PubMed=10782927;
 RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
 RA Fodde R., Alman B., Bapat B.;
 RA "A germline mutation at the extreme 3-prime end of the APC gene
 RT results in a severe desmoid phenotype and is associated with
 RT overexpression of beta-catenin in the desmoid tumor.";
 RL Clin. Genet. 57:205-212(2000).
 RN [7]
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 2-55.
 RP MEDLINE=20384842; PubMed=10926498;
 RA Day C.L., Alber T.;
 RA "Crystal structure of the amino-terminal coiled-coil domain of the
 RT APC tumor suppressor.";
 RL J. Mol. Biol. 301:147-156(2000).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 1021-1035 IN COMPLEX WITH
 CTNNB1.
 RP MEDLINE=21564054; PubMed=11707392;
 RA Eklof Spink K., Fridman S.G., Weis W.I.;
 RA "Molecular mechanisms of beta-catenin recognition by adenomatous
 RT polyposis coli revealed by the structure of an APC-beta-catenin
 RT complex.";
 RL EMBO J. 20:6203-6212(2001).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
 AXIN.
 RP MEDLINE=20271867; PubMed=10811618;
 RA Spink K.E., Polakis P., Weis W.I.;
 RA "Structural basis of the axin-adenomatous polyposis coli
 RT interaction.";
 RL EMBO J. 19:2270-2279(2000).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=94154728; PubMed=9111410;
 RA Nagase H., Nakamura Y.;
 RA "Mutations of the APC (adenomatous polyposis coli) gene.";
 RT Hum. Mutat. 2:425-434(1993).
 RN [11]
 RP VARIANTS FAP.
 RX MEDLINE=91355211; PubMed=1651563;
 RA Nishino I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
 RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
 RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
 RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
 RA "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
 RT patients.";
 RL Science 253:665-669(1991).
 RN [12]
 RP VARIANTS FAP.
 RX MEDLINE=93265030; PubMed=1338904;
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
 RA Miki Y., Mori T., Nakamura Y.;
 RA "Somatic mutations of the APC gene in colorectal tumors: mutation
 RT cluster region in the APC gene.";
 RL Hum. Mol. Genet. 1:229-233(1992).
 RN [13]
 RP VARIANTS FAP.
 RX MEDLINE=93244793; PubMed=1338691;
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
 RA Nakamura Y., Horii A.;
 RA "Somatic mutation of the APC gene in gastric cancer: frequent
 RT mutations in very well differentiated adenocarcinoma and signet-ring
 RT cell carcinoma.";
 RL Hum. Mol. Genet. 1:559-563(1992).
 RN [14]
 RP VARIANT FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND
 RP SER-2502
 RX MEDLINE=93250848; PubMed=1338764;
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,

RA Baba S., Nakamura Y.;
 RT "Screening for germ-line mutations in familial adenomatous polyposis
 RL patients: 61 new patients and a summary of 150 unrelated patients.";
 RL Hum. Mutat. 1:467-473(1992).
 RN [15]
 RP VARIANT FAP TRP-99.
 RC TISSUE=Peripheral blood lymphocytes;
 RX MEDLINE=95134544; PubMed=7831149;
 RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,
 RA Mueller A., Mueller H., Scott R.J.;
 RA "Mutational analysis of the first 14 exons of the adenomatous
 RT polyposis coli (APC) gene.";
 RL Eur. J. Cancer 30A:1709-1713(1994).
 RN [16]
 RP VARIANT FAP GLY-722.
 RX MEDLINE=9513430; PubMed=7833931;
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pillia S., Prete F., Mareni C., Guanti G.;
 RA "Four novel mutations of the APC (adenomatous polyposis coli) gene in
 RT FAP patients.";
 RL Hum. Mol. Genet. 3:1687-1688(1994).
 RN [17]
 RP ERRATUM.
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pillia S., Prete F., Mareni C., Guanti G.;
 RL Hum. Mol. Genet. 3:1918-1918(1994).
 RN [18]
 RP VARIANT FAP ILE-171.
 RX MEDLINE=97144176; PubMed=8990002;
 RA van der Luijt R.B., Meera Khan P., Vasen H.F.A., Tops C.M.J.,
 RA van Leeuwen-Corrielese I.S.J., Wijnen J.T., van der Klift H.M.,
 RA Plug R.J., Griffioen G., Fodde R.;
 RA "Molecular analysis of the APC gene in 105 Dutch kindreds with
 RT familial adenomatous polyposis: 67 germline mutations identified by
 RT DGGE, PTT, and southern analysis.";
 RL Hum. Mutat. 9:7-16(1997).
 RN [19]
 RP VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.
 RX MEDLINE=98080146; PubMed=9419979;
 RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
 RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
 RA Fukutome A., Tomiyama J., Chuganji Y., Momoi M., Utsunomiya J.;
 RA "Drastic genetic instability of tumors and normal tissues in Turcot
 RT syndrome.";
 RL Oncogene 15:2877-2881(1997).
 RN [20]
 RP VARIANT LYS-1307.
 RX MEDLINE=98400248; PubMed=9731522;
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
 RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
 RA "The APC I1307K allele and breast cancer risk.";
 RL Nat. Genet. 20:13-14(1998).
 RN [21]
 RP VARIANTS LYS-1307 AND GLN-1317.
 RC TISSUE=Peripheral blood;
 RX MEDLINE=98393712; PubMed=9724771;
 RA Frayling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
 RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
 RA "The APC variants I1307K and E1317Q are associated with colorectal
 RT tumors, but not always with a family history.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
 RN [22]
 RP VARIANT LYS-1307.
 RX MEDLINE=98400259; PubMed=9731533;
 RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
 RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
 RA "The APC I1307K allele and cancer risk in a community-based study of
 RT Ashkenazi Jews.";
 RL Nat. Genet. 20:62-65(1998).

Query Match				99.7%; Score 14533; DB 1; Length 2843;			
Best Local Similarity				99.8%; Pred No. 0;			
Matches 2836; Conservative				2; Mismatches 5; Indels 0; Gaps 0;			
Qy	1	MAAASYDOLLKOVEALKWENSNLROLEDNENHNLTKLETEASNKEVLUKQOGSIEDEAM	60				
Db	1	MAAASYDOLLKOVEALKWENSNLROLEDNENHNLTKLETEASNKEVLUKQOGSIEDEAM	60				
Qy	61	ASSGOIDLLELKEINLNDSSNPPGVKLSKMSLRSGYSGREGSVASRSCECPVPMGPFPR	120				
Db	61	ASSGOIDLLELKEINLNDSSNPPGVKLSKMSLRSGYSGREGSVASRSCECPVPMGPFPR	120				
Qy	121	RGVNGSRSTGYLBELEKERSLLADLDKEEKEXDWTYAOQLNLTGKIDSLPTENPSL	180				
Db	121	RGVNGSRSTGYLBELEKERSLLADLDKEEKEXDWTYAOQLNLTGKIDSLPTENPSL	180				
Qy	181	QTDLTROLEYEAQIRVAMEBOLCTCODMEKRAQRTARCOOHEKDLIRLOLLQSOAT	240				
Db	181	QTDLTROLEYEAQIRVAMEBOLCTCODMEKRAQRTARCOOHEKDLIRLOLLQSOAT	240				
Qy	241	EAERSQNHETGSHDAERQNEGVGEINMATSGNGQSTTRMDHETASVLSSSSTHSA	300				
Db	241	EAERSQNHETGSHDAERQNEGVGEINMATSGNGQSTTRMDHETASVLSSSSTHSA	300				
Qy	301	PRRLTSHLGTKVEMYVLSLMLGTHDKDDMSHTLLAMSSQDSCISMRQSGCLPLLIQLL	360				
Db	301	PRRLTSHLGTKVEMYVLSLMLGTHDKDDMSHTLLAMSSQDSCISMRQSGCLPLLIQLL	360				
Qy	361	HGNDKDSVLLGNSRGSKEARASAAALHNIHSQPDCKRGREIRVHLLEQIRAYCETC	420				
Db	361	HGNDKDSVLLGNSRGSKEARASAAALHNIHSQPDCKRGREIRVHLLEQIRAYCETC	420				
Qy	421	WEMQEAHEFGMDQKNPMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGGLOIAEILLQ	480				
Db	421	WEMQEAHEFGMDQKNPMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGGLOIAEILLQ	480				
Qy	481	VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCWRALVAQLKSESEDL	540				
Db	481	VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCWRALVAQLKSESEDL	540				
Qy	541	QOVIASVLNLSWRADVNSKKTILREYGSVKALMECALEYKKESTLKSVLNLSAHT	600				
Db	541	QOVIASVLNLSWRADVNSKKTILREYGSVKALMECALEYKKESTLKSVLNLSAHT	600				
Qy	601	ENKADI CAVDGALAFLVGLTLYRSQNTLAIIESGGGILRNYSLLIATNEDHROILRENN	660				
Db	601	ENKADI CAVDGALAFLVGLTLYRSQNTLAIIESGGGILRNYSLLIATNEDHROILRENN	660				
Qy	661	CIQTLLOHLKSHSLTIVSNACGTLNLSARNPKQOEALWDMGAVSMKLNLIHSHKHMIA	720				
Db	661	CIQTLLOHLKSHSLTIVSNACGTLNLSARNPKQOEALWDMGAVSMKLNLIHSHKHMIA	720				
Qy	721	GSAALRNLMANRPAYKXDNIMSPGSSILPSLHVRKQKALEAEALDAQHLSFTFDNIDNLS	780				
Db	721	GSAALRNLMANRPAYKXDNIMSPGSSILPSLHVRKQKALEAEALDAQHLSFTFDNIDNLS	780				
Qy	781	PKASHSKORHKOSLYGVDFDNRHDDNRSDFNTGNMTVLSPYINTVLTPSSSSSRGS	840				
Db	781	PKASHSKORHKOSLYGVDFDNRHDDNRSDFNTGNMTVLSPYINTVLTPSSSSSRGS	840				
Qy	841	LDSSRSSEKORSLERGIGLGNYPHATENPGTSSKRGLOISTTAAQIAKMEEVSAIHTS	900				
Db	841	LDSSRSSEKORSLERGIGLGNYPHATENPGTSSKRGLOISTTAAQIAKMEEVSAIHTS	900				
Qy	901	QEDRSSTTELHCVTDERNALRPSNAATHSNYNTFTKSENRCNCPYAKLEYKRSS	960				
Db	901	QEDRSSTTELHCVTDERNALRPSNAATHSNYNTFTKSENRCNCPYAKLEYKRSS	960				
Qy	961	NDLSINSVSSNDGKGKQMKFPIESYSEDDSEKFCYGYVPADLAHKIHSANHMDNDGE	1020				
Db	961	NDLSINSVSSNDGKGKQMKFPIESYSEDDSEKFCYGYVPADLAHKIHSANHMDNDGE	1020				

Qy	1021	LDTPINYSKYSDQNLNQRQSPQSNRWARPKHIIIEDEIKOSEQRQSRNQSTTYPVYTE	1080				
Db	1021	LDTPINYSKYSDQNLNQRQSPQSNRWARPKHIIIEDEIKOSEQRQSRNQSTTYPVYTE	1080				
Qy	1081	STDDKHLKPOHPGQOECVSPYBSRGANGSETNRVGSNGHGINCVSOSLCOEDDYDDKP	1140				
Db	1081	STDDKHLKPOHPGQOECVSPYBSRGANGSETNRVGSNGHGINCVSOSLCOEDDYDDKP	1140				
Qy	1141	TNYSERYSEEOHEEERPTNYSIKYNEEKRRHVDQPIDYSLKYATDIPSSQKQSFPSKS	1200				
Db	1141	TNYSERYSEEOHEEERPTNYSIKYNEEKRRHVDQPIDYSLKYATDIPSSQKQSFPSKS	1200				
Qy	1201	SSQSSKTEHMSSESSENTSTPSSNAKRONOLHPSAQSRSQPOKAATCKVSSINQETIQ	1260				
Db	1201	SSQSSKTEHMSSESSENTSTPSSNAKRONOLHPSAQSRSQPOKAATCKVSSINQETIQ	1260				
Qy	1261	TYCVEDTPICFSRCSSLSLSSAEDIEGICNOTTOEADSNANTLQIAEIKIGTRSAEDPV	1320				
Db	1261	TYCVEDTPICFSRCSSLSLSSAEDIEGICNOTTOEADSNANTLQIAEIKIGTRSAEDPV	1320				
Qy	1321	SEVPAYVQHPRTKSRLOGSSLSSEBARHKAFFPSGAKSPSKGAOTPKSPPEHYVQET	1380				
Db	1321	SEVPAYVQHPRTKSRLOGSSLSSEBARHKAFFPSGAKSPSKGAOTPKSPPEHYVQET	1380				
Qy	1381	PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGQTMPSPRSKTPP	1440				
Db	1381	PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGQTMPSPRSKTPP	1440				
Qy	1441	PPPQTAQTKREVPKAKPTAEKESGPKAAVNAQVORVQVLPDADTLHLHFAESTPDGF	1500				
Db	1441	PPPQTAQTKREVPKAKPTAEKESGPKAAVNAQVORVQVLPDADTLHLHFAESTPDGF	1500				
Qy	1501	SCSSLSALSILDEPFIQKVELRIMPVQENDNGNETSEOPKESNENQKEAEKTIIDSE	1560				
Db	1501	SCSSLSALSILDEPFIQKVELRIMPVQENDNGNETSEOPKESNENQKEAEKTIIDSE	1560				
Qy	1561	KDLLDDDDDDIEILEECIISAMPTKSSRKGKPAQATASKLPPPPVARKPSQLPVYKLLPS	1620				
Db	1561	KDLLDDDDDDIEILEECIISAMPTKSSRKGKPAQATASKLPPPPVARKPSQLPVYKLLPS	1620				
Qy	1621	QNRLOPKHVSFTPGDDMPRYCVGEGTPIINFSTATSLSDLTISPPNELAAGEVRGGAQ	1680				
Db	1621	QNRLOPKHVSFTPGDDMPRYCVGEGTPIINFSTATSLSDLTISPPNELAAGEVRGGAQ	1680				
Qy	1681	SBEFEKRTTPTGSRSTDEAGGKTSVTIPELDDNKAEEGDILAEICINAMPKGSKHP	1740				
Db	1681	SBEFEKRTTPTGSRSTDEAGGKTSVTIPELDDNKAEEGDILAEICINAMPKGSKHP	1740				
Qy	1741	PRVKIMQVOQASASSAPNKNQLDGKKKPTSPVPKIPONTETRYTRVRKNADSNNLN	1800				
Db	1741	PRVKIMQVOQASASSAPNKNQLDGKKKPTSPVPKIPONTETRYTRVRKNADSNNLN	1800				
Qy	1801	AERFVSNKSKONLNKNSKDFNDKLPNNEDRVGSGFADSPHHTPIEGTPTCFGRND	1860				
Db	1801	AERFVSNKSKONLNKNSKDFNDKLPNNEDRVGSGFADSPHHTPIEGTPTCFGRND	1860				
Qy	1861	SLSLDFDDDDVDLSREKAEALRKAENKESFAKVTSTHETSTNQQSANKTQAIKQPINR	1920				
Db	1861	SLSLDFDDDDVDLSREKAEALRKAENKESFAKVTSTHETSTNQQSANKTQAIKQPINR	1920				
Qy	1921	QGPXILQKQSTFPQSSKQIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDENN	1980				
Db	1921	QGPXILQKQSTFPQSSKQIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDENN	1980				
Qy	1981	NKENPIKETETPPPSQSGEPSPKQASGYAPKSFHVHVEDTFCFSRNSLSLSLSDIDELLQ	2040				
Db	1981	NKENPIKETETPPPSQSGEPSPKQASGYAPKSFHVHVEDTFCFSRNSLSLSLSDIDELLQ	2040				
Qy	2041	ECISSAMPKKKPSRLKDNKXHSFRNNGGILGEDLTLDKDIQRPDSEHGLSPDSENF	2100				
Db	2041	ECISSAMPKKKPSRLKDNKXHSFRNNGGILGEDLTLDKDIQRPDSEHGLSPDSENF	2100				
Qy	2101	WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSILSLKSGISLSPHLPDQBEKEFT	2160				

Db 2101 WKAQEGANSVSSLHQAAAAACLSRQASSSDLSLSKSGISGSPFHLTPDGEKPF 2160
 QY 2161 SNKGPRILKPKGEKSTLETKKIESSEKGIKGGKVKYKSLITCKVRSNBEISQMKQKQLOAN 2220
 Db 2161 SNKGPRILKPKGEKSTLETKKIESSEKGIKGGKVKYKSLITCKVRSNBEISQMKQKQLOAN 2220
 QY 2221 MPSISRGRTMIHICPVNRSSSTSPVSKKGPPLKTPASKSEGCQTATTSPRGAKPVSKS 2280
 Db 2221 MPSISRGRTMIHICPVNRSSSTSPVSKKGPPLKTPASKSEGCQTATTSPRGAKPVSKS 2280
 QY 2281 ELSPVARTQSIGSSKAPSPSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Db 2281 ELSPVARTQSIGSSKAPSPSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
 QY 2341 KLSQLPRTSSPTASTKSSGSKMSYSPGRMSQOQLTKOTGLSKVASSIPRSESASKG 2400
 Db 2341 KLSQLPRTSSPTASTKSSGSKMSYSPGRMSQOQLTKOTGLSKVASSIPRSESASKG 2400
 QY 2401 LNQMNGNGANKVELSRMSSTKSSGESDRSERPVLVROSTFIKEAPSPTLRKLKLESA 2460
 Db 2401 LNQMNGNGANKVELSRMSSTKSSGESDRSERPVLVROSTFIKEAPSPTLRKLKLESA 2460
 QY 2461 SFESLSPSRPASPTRSQAQTPVLSPSLPMSLSTHSSVQAGGWRKLPPLNLSPTIEYNDG 2520
 Db 2461 SFESLSPSRPASPTRSQAQTPVLSPSLPMSLSTHSSVQAGGWRKLPPLNLSPTIEYNDG 2520
 QY 2521 RPAKHDIARSHSPSRPLNRSCTWKREHSKSSSLPRVSTWRTGSSSSILSASSES 2580
 Db 2521 RPAKHDIARSHSPSRPLNRSCTWKREHSKSSSLPRVSTWRTGSSSSILSASSES 2580
 QY 2581 SEKAKSEKHNISIGTKSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 Db 2581 SEKAKSEKHNISIGTKSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 QY 2641 KTLIQVAPAVSKTDEYVWRTEPCIPNPSGRSPGTCNTPPVDSVSEKANPNIKOSKON 2700
 Db 2641 KTLIQVAPAVSKTDEYVWRTEPCIPNPSGRSPGTCNTPPVDSVSEKANPNIKOSKON 2700
 QY 2701 QAKQNVGNGVPMRTVGLNRLNFIQVDAPDQKTEIKPGQNNPVVSETNESPIVERT 2760
 Db 2701 QAKQNVGNGVPMRTVGLNRLNFIQVDAPDQKTEIKPGQNNPVVSETNESPIVERT 2760
 QY 2761 PFSSSSSKHSPSGTVAARTVPPNPNPSPRKSSADTSARPQIPTPVNNNTKKRDSKT 2820
 Db 2761 PFSSSSSKHSPSGTVAARTVPPNPNPSPRKSSADTSARPQIPTPVNNNTKKRDSKT 2820
 QY 2821 DSTESSGTQSPKRSYGLVTSV 2843
 Db 2821 DSTESSGTQSPKRSYGLVTSV 2843

RESULT 2
 APC_RAT ID - APC RAT STANDARD; PRT; 2842 AA.
 AC P70478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
 RA Sugimura T., Nagao M.;
 RT "cDNA cloning of the rat APC gene and assignment to chromosome 18";
 PL Mamm. Genome 6:746-748(1995).
 Query Match 90.3%; Score 13168.5; DB 1; Length 2842;
 Best Local Similarity 89.8%; Pred. No. 0;
 Matches 2557; Conservative 123; Mismatches 158; Indels 9; Gaps 8;
 QY 1 MAASVDQLLKQVALKMENSINLROELEDNSNHLTKLETEASNKKEVLKOLQGSIEDEAM 60
 Db 1 MAASVDQLLKQVALKMENSINLROELEDNSNHLTKLETEASNKKEVLKOLQGSIEDEAM 60
 QY 61 ASSGQIDLLERLKEINLDSSNFFPGVKLSKMSLSYSGSRGVSRSRSGECSVPVPMGSPFR 120
 Db 61 -TSGQIDLLERLKEFNLD-SNFPGVKLSKMSLSYSGSRGVSRSRSGECSVPVPMGSPFR 118
 QY 121 RGFVNGSRSTGYLEELKERSILLADLKEEKKDWYAAQLQNLTKRISLPTENFSL 180
 Db 119 RAFVNGSRSTGYLEELKERSILLADLKEEKKDWYAAQLQNLTKRISLPTENFSL 178
 QY 181 QTDLTRQLEYEARQIRVAMEEQGTQDMKRAQRRIARIQOKEKILIRQLQSOAT 240
 Db 179 QTDLTRQLEYEARQIRVAMEEQGTQDMKRAQRRIARIQOKEKILIRVRLQSOAA 238
 QY 241 BAERSSQNKHTGSHDAERQNEGGVGEINMATSNGCGSTTRMDHETASVLSSSSTHSA 300

RP MUTAGENESIS.
 RC STRAIN=Sprague-Dawley, and Fischer 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Nagao M.,
 RA Weisburger J.H., Sugimura T., Nagao M.;
 RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
 tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914 (1995).
 CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 and participates in Wnt signaling. APC activity is correlated with
 its phosphorylation state (By similarity).
 CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
 axin (By similarity).
 CC -!- PTM: Phosphorylated by GSK3B (By similarity).
 CC -!- SIMILARITY: Contains 7 ARM repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D38629; BAA07609.1; -.
 DR HSP; Q02248; 3BCT.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg; 4.
 DR SMART; SM00185; ARM; 5.
 DR PROSITE; PS0176; ARM_REPEAT; 1.
 KW Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
 KW Repeat.
 FT DOMAIN 1 728 LEU-RICH.
 FT DOMAIN 1 62 COILED COIL (POTENTIAL).
 FT DOMAIN 125 260 COILED COIL (POTENTIAL).
 FT REPEAT 451 493 ARM 1.
 FT REPEAT 503 545 ARM 2.
 FT REPEAT 546 589 ARM 3.
 FT REPEAT 590 636 ARM 4.
 FT REPEAT 637 681 ARM 5.
 FT REPEAT 682 723 ARM 6.
 FT REPEAT 724 765 ARM 7.
 FT DOMAIN 739 2831 SER-RICH.
 FT DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 HIGHLY CHARGED.
 FT MUTAGEN C->R; IN AN IQ-INDUCED COLON TUMOR.
 SQ SEQUENCE 2842 AA; 310530 MW; 3CB2EA8A34E8F47 CRC64;

Db 239 EAERSSQSKHETASHEARQLEGQGVAEISNATSGSQSSAARVDHETAGVLSSGTHSA 298
Qy 301 PRRLTSHLGTKEVYISLLSMLGTHDKDMSRTLLAWSSQSDSCISMQRQGCPLLIQLL 360
Db 299 PRRLTSHLGTKEVYISLLSMLGTHDKDMSRTLLAWSSQSDSCISMQRQGCPLLIQLL 358
Qy 361 HGNDKDSVLLGNSRGSKEARASAAALNIIHSDPPDKRREIRVLLHLEIQIRAYCETC 420
Db 359 HGNDKDSVLLGNSRGSKEARASAAALNIIHSDPPDKRREIRVLLHLEIQIRAYCETC 418
Qy 421 WEMQEAHEPGMDQDNPMFAVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAIAELLQ 480
Db 419 WEMQEAHEPGMDQDNPMFAVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAIAELLQ 478
Qy 481 VDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSKMGCMRALVAQLKSESEDL 540
Db 479 VDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSKMGCMRALVAQLKSESEDL 538
Qy 541 QOVVASVLRLNLSWRADVNSKTLREVGSKALMECALVEKESITLKSVLKALMNLASHT 600
Db 539 QOVVASVLRLNLSWRADVNSKTLREVGSKALMECALVEKESITLKSVLKALMNLASHT 598
Qy 601 ENKADICAVDGAFLVGLTYRSGTNTLAIIESGGGILRVNSVSLIATNEDHROILRENN 660
Db 599 ENKADICAVDGAFLVGLTYRSGTNTLAIIESGGGILRVNSVSLIATNEDHROILRENN 658
Qy 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARPKEOALMDWGMVLMKLNLIHSHKXIAM 720
Db 659 CLQTLLOHLKSHSLTIVSNACGTLNLSARPKEOALMDWGMVLMKLNLIHSHKXIAM 718
Qy 721 GSAALRLNLMANRPKAKYKDNIMSPGSLPSLHVRKQKALEAELDAQHLSFTFNIDNLS 780
Db 719 GSAALRLNLMANRPKAKYKDNIMSPGSLPSLHVRKQKALEAELDAQHLSFTFNIDNLS 778
Qy 781 PKASHRSKORHQSILYGDVFTNHRDNRDNFTGNMTVLSPLYNTTIVLPSSSSRGS 840
Db 779 PKASHRSKORHQSILYGDVFTNHRDNRDNFTGNMTVLSPLYNTTIVLPSSSSRGS 838
Qy 841 LSSSEKDRSLERERG:GLGNHYHATENPGTSSKRGILQISTTAAQIAKWEEVSALHTS 900
Db 839 LSSSEKDRSLERERG:GLSYHATENPGTSSKRGILQISATAAQIAKWEEVSALHTS 898
Qy 901 QEDRSSGTTLHCVTDERNALRRSAATHNTVNTFKSENSNPTCMYPKALEYKRSS 960
Db 899 QEDRSSGTTLHCVTDERNALRRSAATHNTVNTFKSENSNPTCMYPKALEYKRSS 958
Qy 961 NDLSNVSNDGCKRGQMKPSIBSYSEDESKFCYGOYPADLAHKIHSANHMDNDGE 1020
Db 959 NDLSNVSNDGCKRGQMKPSIBSYSEDESKFCYGOYPADLAHKIHSANHMDNDGE 1018
Qy 1021 LPTPINYSILKYDEQLNSGRQSPQNERWARPKHIEDEIKQSEQRQSRNQSTTVPVYTE 1080
Db 1019 LPTPINYSILKYDEQLNSGRQSPQNERWARPKHIEDEIKQSEQRQSRNQSTTVPVYTE 1078
Qy 1081 STDDXHLKFPQHFQOECVSPYRSGANGSEFTNRVGNHGINQVYSQILCOEDDYEDDKP 1140
Db 1079 NTDDXHLKFPQHFQOECVSPYRSGANGSEFTNRVGNHGINQVYSQILCOEDDYEDDKP 1138
Qy 1141 TNYSEYSEEEHEBERPTNYSIKYNEKHVDQPIDYSLKATDIPSSQKQSFSSKS 1200
Db 1139 TNYSEYSEEEHEBERPTNYSIKYNEKHVDQPIDYSLKATDIPSSQKQSFSSKS 1198
Qy 1201 SSCQSKSTHEMSSSNTSTPSNAKRONQLHPSSAQSRSGOPQAACTCKVSSINQETIQ 1260
Db 1199 PSYQGTKEHNSPSSAASAPSNKROSLHPSSAQ-RNGQTPKGTACKVPSINQETMQ 1257
Qy 1261 TYCVEDTPICFRCSLSLSAEDRIGCNQTTQEADSAANTLQIAEIKGIGTRGAEDPV 1320
Db 1258 TYCVEDTPICFRCSLSLSAEDRIGCNQTTQEADSAANTLQIAEIKGIGTRGAEDPV 1317
Qy 1321 SEYPAVSQHPRTKSSRQSSLSSESARHAKVEFPSSGAKSPSKGAQTTPKSPPEHYVQET 1380
Db 1318 SDYPAVSQHPRTKSSRQSSLSSESARHAKVEFPSSGAKSPSKGAQTTPKSPPEHYVQET 1377

Qy 1381 PLMFSECTSVSSLDSEFBSRSIASSVQSEFSCSGMVSGIISPSDLPSPGQTMPPSSSKT-- 1438
Db 1378 PLVFSECTSVSSLDSEFBSRSIASSVQSEFSCSGMVSGIISPSDLPSPGQTMPPSSSKT 1437
Qy 1439 PPPPPPTAGTKKEVPKQKAPTAEKRESGPKQAANAAVQVQVLPDADTLLHFATESTPD 1498
Db 1438 PPPPPPTAGTKKEVPKQKAPTAEKRESGPKQAANAAVQVQVLPDADTLLHFATESTPD 1497
Qy 1499 GFSCSSSLSALSDEPFIQKVELRIMPVQENDNGNETESEPQKESNENQEKAKTID 1558
Db 1498 GFSCSSSLSALSDEPFIQKVELRIMPVQENDNGNETESEPQKESNENQEKAKTID 1556
Qy 1559 SEKDLLDDDDDDIEILEECIISAMPTKSSRKPKQAQATASKLPPPVARKPSOLPVYKLL 1618
Db 1557 SEKDLLDDDDDDIEILEECIISAMPTKSSRKPKQAQATASKLPPPVARKPSOLPVYKLL 1616
Qy 1619 PSQNRLOPQKHVSFTFGDDMPRVYCVETGPIINFSTATISLDTIESPPNELAAGEVRGG 1678
Db 1617 PSQNRLOPQKHVSFTFGDDMPRVYCVETGPIINFSTATISLDTIESPPNELAAGEVRGG 1676
Qy 1679 AQSGBEFKDDTITEGRSTDEACGGKTSVVTPELDNKAEEGDIILAEICINSAMPKXSH 1738
Db 1677 VQSGEFKDDTITEGRSTDEACGGKTSVVTPELDNKAEEGDIILAEICINSAMPKXSH 1736
Qy 1739 KPRVRVKIMDQVQOASASASAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKN 1798
Db 1737 KPRVRVKIMDQVQOASASASAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKN 1796
Qy 1799 LNAERVFSDKSKQKQNLKNNKDFNDKLPNNEDVRGSAFDSPHYPIEGTPTCESR 1858
Db 1797 VNTTEFTSDKSKQKQNLKNNKDFNDKLPNNEDVRGSAFDSPHYPIEGTPTCESR 1856
Qy 1859 NDLSLSDLDVDSREKAEELKAKENKESAKVTSHTLTSTNOQSANKTQAIKQPI 1918
Db 1857 NDLSLSDLDVDSREKAEELKAKENKESAKVTSHTLTSTNOQSANKTQAIKQPI 1916
Qy 1919 NRQOPKILQKSTFQSSKDIIDRGAATDEKQNFATIENTPVCFSHNSLSLSLSDOE 1978
Db 1917 NRQOPKILQKSTFQSSKDIIDRGAATDEKQNFATIENTPVCFSHNSLSLSLSDOE 1976
Qy 1979 -NNKENEPKETPEPDSQGEPSKPOASGVAPKSFVEDTPVCFSRNSLSLSLSDIEDD 2037
Db 1977 NNNEETGVRDAPANAQOQKQKQASGVAPKSFVEDTPVCFSRNSLSLSLSDIEDD 2036
Qy 2038 LLOECISSAMPKPKKPSRLKGDNEKHSRNMNGIIGEDLFLDKDQRPDSEHGLSPDSE 2097
Db 2037 LLRECISSAMPKPKKPSRLKGDNEKHSRNMNGIIGEDLFLDKDQRPDSEHGLSPDSE 2096
Qy 2098 NFDWKATQEGANSTVSSLIHQ-AAAAACLSQASDSDSILSLKSGISLGSPEHLTPDOEE 2156
Db 2097 NFDWKATQEGANSTVSSLIHQ-AAAAACLSQASDSDSILSLKSGISLGSPEHLTPDOEE 2156
Qy 2157 KPFTSNKGPRLKPGKSTLTETKKIESKGIKGGKVKYKSLITGKVRNSEISGQMKQP 2216
Db 2157 KPFTSNKGPRLKPGKSTLTETKKIESKGIKGGKVKYKSLITGKVRNSEISGQMKQP 2216
Qy 2217 LOANPISIRGRWTMIHPGVNRSSSTSPVSKGPPKLTASKSPBEGQATTPSPGAKP 2276
Db 2217 LOANPISIRGRWTMIHPGVNRSSSTSPVSKGPPKLTASKSPBEGQATTPSPGAKP 2276
Qy 2277 SVKSELSPVARQTSOIGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISPGNGI 2336
Db 2277 AVKSELSPITQTSIHISGSKNGPSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISPGNGI 2336
Qy 2337 SPPNKLSQLPRTSSPSTASTKSGSGKQVYTSPPGRQMSQONLTKQTLGSKVASSIPRSES 2396
Db 2337 STPNKLSQLPRTSSPSTASTKSGSGKQVYTSPPGRQMSQONLTKQTLGSKVASSIPRSES 2396
Qy 2397 ASKGLNOMNNGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTFLRKL 2456
Db 2397 ASKGLNOMNNGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTFLRKL 2456

QY 2457 BESASFESLSRSPASPTRSQAOPTVLPSPSLPMSLSTHSSVOAGGWRKLPNLSPTIE 2516
 DB 2457 BESASFESLSRSPASPTRSQAOPTVLPSPSLPMSLSTHSSVOAGGWRKLPNLSPTIE 2516
 QY 2517 YNDGRPAKHDIARSHSPSRPLPNRSQGTWRKSHSSSLPRVSTWRTGSSSSILSA 2576
 DB 2517 YSDGRPSKHDIARSHSPSRPLPNRSQGTWRKSHSSSLPRVSTWRTGSSSSILSA 2576
 QY 2577 SSESSEKAKSDEKHNVSISGTSKQSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATN 2636
 DB 2577 SSESSEKAKSDEKHNVSISGTSKQSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATN 2636
 QY 2637 GAESKTLIYQAPAVSKTQVWRIEDCPINNPRSGSPGTGNTPPVIDSSEKXANPNKID 2696
 DB 2637 GAESKTLIYQAPAVSKTQVWRIEDCPINNPRSGSPGTGNTPPVIDSSEKXANPNKID 2696
 QY 2697 SKDQAQKQNVGSPVMTVGLNRLTSFIQVADPDKQTEIKPGQNNPVVSETNESPI 2756
 DB 2697 SKDQAQKQNVGSPVMTVGLNRLTSFIQVADPDKQTEIKPGQNNPVVSETNESPI 2756
 QY 2757 VERTPFSSSSSKHSSPSGIVAAVTPFNPNYSPRKSADSTGAPSPQIPTPVNNNTKKR 2816
 DB 2756 AERTPFSSSSSKHSSPSGIVAAVTPFNPNYSPRKSADSTGAPSPQIPTPVNNNTKKR 2816
 QY 2817 DSKTDSSTESSGTQSPKHSGLSYLTVSV 2843
 DB 2816 DSKTDSSTESSGTQSPKHSGLSYLTVSV 2842

RESULT 3
 ID APC MOUSE
 AC Q61315, Q62044; STANDARD; PRT: 2845 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein) (mAPC).
 GN APC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
 RC STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;
 RX MEDLINE=92263101; PubMed=1350108;
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.;
 RA "Multiple intestinal neoplasia caused by a mutation in the murine
 RT homolog of the APC gene."
 RL Science 256:668-670 (1992).
 RN [2]
 RP ERRATUM.
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.;
 RL Science 256:1114-1114 (1992).
 RN [3]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Dicker F., Lambertz S., Reitmair A., Ballhausen W.G.;
 RT "The murine APC gene: alternative splicing of 5' untranslated
 RL region segments."
 RN Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=94061824; PubMed=8242607;
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
 RT "APC gene messenger RNA: novel isoforms that lack exon 7";
 RL Cancer Res. 53:5589-5591 (1993).
 CC -i- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 CC and participates in Wnt signaling. APC activity is correlated with
 CC its phosphorylation state (By similarity).
 CC -i- SUBUNIT: Forms homooligomers. Associates with catenins. Binds

CC axin (By similarity).
 CC -i- ALTERNATIVE PRODUCTS.
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=1;
 CC IsoId=Q61315-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q61315-2; Sequence=VSP_0041116;
 CC Name=3;
 CC IsoId=Q61315-3; Sequence=VSP_0041117;
 CC Name=4;
 CC IsoId=Q61315-4; Sequence=VSP_0041116, VSP_0041117;
 CC -i- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
 CC lung, brain, stomach, intestine, testis and ovary.
 CC -i- PTM: Phosphorylated by GSK3B (By similarity).
 CC -i- SIMILARITY: Contains 7 ARM repeats.
 CC This SWISS-PROT entry is copy-right. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M8127; AAB59632.1; -;
 CC EMBL; U02927; AAB3443.1; -;
 CC FIR; I49505; I49505.
 CC HSP; Q02248; 3BCT.
 CC MGD; MGI:88039; Apc.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC GO; GO:0008013; F:beta-catenin binding; IDA.
 CC GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
 CC GO; GO:0009953; P:axis specification; IMP.
 CC GO; GO:0016055; P:dorsal/ventral pattern formation; IMP.
 CC InterPro; IPR008938; ARM.
 CC InterPro; IPR000225; Armadillo.
 CC Pfam; PF00514; Armadillo_seg; 4.
 CC SMART; SM00185; ARM; 5 REPEAT; 1.
 CC PROSITE; PS0176; ARM_REPEAT; 1.
 CC Wnt signaling pathway; Anti-oncogene; Phosphorylation;
 CC Alternative splicing; Repeat; Coiled coil.
 CC DOMAIN 1 61
 CC COILED COIL (POTENTIAL).
 CC COILED COIL (POTENTIAL).
 CC LEU-RICH.
 CC ARM 1.
 CC ARM 2.
 CC ARM 3.
 CC ARM 4.
 CC ARM 5.
 CC ARM 6.
 CC ARM 7.
 CC SER-RICH.
 CC ASP/GLU-RICH (ACIDIC).
 CC ASP/GLU-RICH (ACIDIC).
 CC HIGHLY CHARGED.
 CC Missing (in isoform 2 and isoform 4).
 CC /FTid=VSP_0041116.
 CC Missing (in isoform 3 and isoform 4).
 CC /FTid=VSP_0041117.
 CC T -> A (IN STRAIN CAST/EI).
 CC V -> I (IN STRAIN CAST/EI).
 CC Y -> F (IN STRAIN CAST/EI).
 CC A -> T (IN STRAIN CAST/EI).
 CC G -> S (IN STRAIN CAST/EI).
 CC G -> A (IN STRAIN CAST/EI).
 CC H -> Q (IN STRAIN CAST/EI).
 CC T -> A (IN STRAIN CAST/EI).
 CC T -> S (IN STRAIN CAST/EI).
 CC SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A99 CRC64;
 CC 90.2%; Score 13148; DB 1; Length 2845;
 CC Query Match

Best Local Similarity 90.0%; Pred. No. 0;
Matches 2566; Conservative 113; Mismatches 159; Indels 12; Gaps 10;

QY 1 MAASDYDOLLKQVAKWENLNKQELLEDNSNHLTKLETEASNKKEVLKQLOSGIEDAM 60
DB 1 MAASDYDOLLKQVAKWENLNKQELLEDNSNHLTKLETEASNKKEVLKQLOSGIEDAM 60
QY 61 ASSGOIDLLELKEINLDSNFFPGYKLRKMSLRSYSGREGSVSSRGEGCSPVPMGSPFR 120
DB 61 -TSQIDLLERLKEFNLD-SNFGYKLRKMSLRSYSGREGSVSSRGEGCSPVPMGSPFR 118
QY 121 RGFVNGSRSTGYLEELKERSLLADLDKEKDWYLAQONLTKRIDSPLTENFSL 180
DB 119 RTFVNGSRSTGYLEELKERSLLADLDKEKDWYLAQONLTKRIDSPLTENFSL 178
QY 181 QTDLTRRQLEYEARQIRVAMEBEQLGTCODMEKRAORRIARQQIEKODILRIROLLOQOAT 240
DB 179 QTDLTRRQLEYEARQIRVAMEBEQLGTCODMEKRAORRIARQQIEKODILRIROLLOQOAA 238
QY 241 EABRSSQNKHTGSHDABRQNEGGQVGEINMATSGNGGGSTRMDHETASVLSSSSTHSA 300
DB 239 EABRSSQNRHDAASHEAGROHEGHAESNTAASSGGSPATRYVDHETASVLSSSGTHSA 298
QY 301 PRLTSHLGTKEVYVLSLVLGTHDKDDMSRTLLAMSSSODSCISMRQSGCLPLLQLL 360
DB 299 PRLTSHLGTKEVYVLSLVLGTHDKDDMSRTLLAMSSSODSCISMRQSGCLPLLQLL 358
QY 361 HGNDKDSVLLNGSRGSEKARASAAALHNIHSPDDKRGREIRVHLLEQIRAYCETC 420
DB 359 HGNDKDSVLLNGSRGSEKARASAAALHNIHSPDDKRGREIRVHLLEQIRAYCETC 418
QY 421 WEMQEAHEPGNDQKNPAPVEHQICPAVCVLMKLSFDEEHRHAMELGLQIAIELLQ 480
DB 419 WEMQEAHEPGNDQKNPAPVEHQICPAVCVLMKLSFDEEHRHAMELGLQIAIELLQ 478
QY 481 VDCMYGLTNDHYSTITLRYAGMALTNLTFGDVANKATLCMKMGCMRALVAQLKSEEDL 540
DB 479 VDCMYGLTNDHYSTITLRYAGMALTNLTFGDVANKATLCMKMGCMRALVAQLKSEEDL 538
QY 541 QOVTASVLRLNLSRADVNSKTLREVGSVKALMECALEVKKESTLKVLSALMNLASHCT 600
DB 539 QOVTASVLRLNLSRADVNSKTLREVGSVKALMECALEVKKESTLKVLSALMNLASHCT 598
QY 601 ENKADICAVDQALAFVGLTVRSQNTLAIIESGGGILRVNSVSLIATNEDHROILRENN 660
DB 599 ENKADICAVDQALAFVGLTVRSQNTLAIIESGGGILRVNSVSLIATNEDHROILRENN 658
QY 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKQOALWDMGAVSMLKNLIHSHKMIAM 720
DB 659 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKQOALWDMGAVSMLKNLIHSHKMIAM 718
QY 721 GSAALRLNLMANRPKAYKDANIMSGSSLPSLHYRKOKALEAELDAOHLSETFNIDNLS 780
DB 719 GSAALRLNLMANRPKAYKDANIMSGSSLPSLHYRKOKALEAELDAOHLSETFNIDNLS 778
QY 781 PKASHRSQRHKSQSLYGDYVDFTRNDRDNRDNFNNTGNMTVLSPLYNTTIVLPSSSSSRGS 840
DB 779 PKASHRSQRHKSQSLYGDYVDFTRNDRDNRDNFNNTGNMTVLSPLYNTTIVLPSSSSSRGS 838
QY 841 LDSSRSKDRSLERBERGIGLGNYPATENPGTSKRGLOISTTAAQIAKVMEEVSAIHTS 900
DB 839 LDSSRSKDRSLERBERGIGLGNYPATENPGTSKRGLOISTTAAQIAKVMEEVSAIHTS 898
QY 901 QEDRSSGGTTELHCVTDERNALRRSSAAHTSNTYNTFKSENSNRTCSMPYAKLEYKRSS 960
DB 899 QEDRSSGGTTELHCVTDERNALRRSSAAHTSNTYNTFKSENSNRTCSMPYAKLEYKRSS 958
QY 961 NDSLNSVSNDDGYKRGQMKPSIESYEDDSKFCSCYQYQYQYQYQYQYQYQYQYQYQY 1020
DB 959 NDSLNSVSNDDGYKRGQMKPSIESYEDDSKFCSCYQYQYQYQYQYQYQYQYQYQYQY 1018
QY 1021 LDTPINYSILKYDEQLNSGRQSPQNERWAPKPHVIEDEIKQNEQORQARSQNTSPYVYSE 1078

DB 1019 LDTPINYSILKYDEQLNSGRQSPQNERWAPKPHVIEDEIKQNEQORQARSQNTSPYVYSE 1078
QY 1081 STDDKHLKXQPHFGQOECVSPYRSGANGSENRVCSNHNINONVSQSLCQDDDEDDEDDP 1140
DB 1079 NTDDKHLKXQPHFGQOECVSPYRSGTSGSETNRMGSHAINONVNQSLCQDDDEDDEDDP 1138
QY 1141 TNYSEYSEEBQH-EHEERTNYSIKYNEEKHVDOPIDYSLKYATDIPSSQKQSFPSK 1199
DB 1139 TNYSEYSEEBQH-EHEERTNYSIKYNEEKHVDOPIDYSLKYATDIPSSQKQSFPSK 1198
QY 1200 SSGQSKTEHWSSESSTENTSPSSNAKRONQHPASAQSRSGOPAKAATCKVSSINQETI 1259
DB 1199 NSQAQTKPEHLSPPSENTAVPPSNAKRONQHPASAQ-RNGCTQGTCTCKVPSINQETI 1257
QY 1260 QTYCVEDTPICFSRCSLSLSAEDEICNCTOTQADANTLQIAEIKGKIGTRSAEDP 1319
DB 1258 QTYCVEDTPICFSRCSLSLSAEDEICNCTOTQADANTLQIAEIKGKIGTRSAEDP 1317
QY 1320 VSEVPVAVSQPRTKSSRLQSSLSSESABH-KAVEPPSGAKSPSKGCAOTPKSPPHYVQ 1378
DB 1318 ATEVPVAVSQPRTKSSRLQSSLSSESABH-KAVEPPSGAKSPSKGCAOTPKSPPHYVQ 1377
QY 1379 ETPLMFSTRCTSVSSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPDSQGTMPSPRSKT 1438
DB 1378 ETPLMFSTRCTSVSSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPDSQGTMPSPRSKT 1437
QY 1439 PPPPOTATQKREVPKXKAPTAKRESGKQAANVAQVQVLPDADTLHLPATSTPD 1498
DB 1438 PPPPOTATQKREVPKXKAPTAKRESGKQAANVAQVQVLPDADTLHLPATSTPD 1497
QY 1499 GFSCSSLSALSDEPFIQKVELRIMPPVQENDNGNETESQPKESNENQOEKAKTID 1558
DB 1498 GFSCSSLSALSDEPFIQKVELRIMPPVQENDNGNETESQPKESNENQOEKAKTID 1556
QY 1559 SEKDLDDDDDDIEILEECIIISAMPTKSRKPKPAQATASKLPPPVARKPSQLPYKLL 1618
DB 1557 SEKDLDDDDDDIEILEECIIISAMPTKSRKPKPAQATASKLPPPVARKPSQLPYKLL 1616
QY 1619 PSQNRLOPKHVSFTPGDDMPVYCVGEGTPIINFSTATSLDITIEBPNELAAGEVRG 1678
DB 1617 PAQNRLOPKHVSFTPGDDMPVYCVGEGTPIINFSTATSLDITIEBPNELAAGEVRG 1676
QY 1679 AQSGEPEKRDITPTGEGSTDEAQGGTSSVTTPPELDNKAEEGDIILAEICINSAMPKXSH 1738
DB 1677 IQSGEPEKRDITPTGEGSTDEAQGGTSSVTTPPELDNKAEEGDIILAEICINSAMPKXSH 1736
QY 1739 KPFRVKIMDQVQOQASASSAPNQNOLDGKXKPTSPVKPIPONTYRTRVRKXNADSKN 1798
DB 1737 KPFRVKIMDQVQOQASASSAPNQNOLDGKXKPTSPVKPIPONTYRTRVRKXNADSKN 1796
QY 1799 LNAERVFSNDKOSKONLKNNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTPTCFSR 1858
DB 1797 VNTETPFSNDKOSKONLKNNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTPTCFSR 1856
QY 1859 NDSLSDLDDEDDVLSREKAEIRKAKENKSEAKVTSHTELTSSNOQANKQIAIAKQPI 1918
DB 1857 NDSLSDLDDEDDVLSREKAEIRKAKENKSEAKVTSHTELTSSNOQANKQIAIAKQPI 1916
QY 1919 NRGOPKILQKSTFPQSSKDIIDRGAATDEKLOKFAIENTPVCFSHNSLSLSLSDIDQE 1978
DB 1917 NRGOPKILQKSTFPQSSKDIIDRGAATDEKLOKFAIENTPVCFSHNSLSLSLSDIDQE 1976
QY 1979 -NNKNEEPIKETEPEDSQGEPSPKQASGYAPKSFHVEDTPVCFSSNSLSLSLSDIDSE 2037
DB 1977 -NNKNEEPIKETEPEDSQGEPSPKQASGYAPKSFHVEDTPVCFSSNSLSLSLSDIDSE 2036
QY 2038 LLOECISSAMPKXKPSRLKGNDEKHSPRNMGGILGEDITLDLKDILQRPDSEHGLSPDE 2097
DB 2037 LLOECISSAMPKXKPSRLKGNDEKHSPRNMGGILGEDITLDLKDILQRPDSEHGLSPDE 2096
QY 2098 NFDWKALQEGANSIVSLHQ-AAAAACLQRQASSDSDSILSKSGISLSPFHLTDPQEE 2156
DB 2097 NFDWKALQEGANSIVSLHQ-AAAAACLQRQASSDSDSILSKSGISLSPFHLTDPQEE 2156

QY	932	SNTYNFTKSEN-----SNRTCSMPYAKLEYKRSS-----NDLSNVSNSNGY-----	973
Db	382	GESSEALKAERRLIKOKAGKHILKEKISKLEEKKOKKEKKEIKERKELKEEGRKEEK	441
QY	974	-----GRRGOMKPSIESYDESDSKFSYQYQYPADLAHKIHSANHMDDNDGELDTPTNYS	1028
Db	442	DAKDXEKKOTKPEVKXLSKPLKPF-----TPEVRKTLYAKAKAPGRVKVDKGRA	491
QY	1029	LKYSDLEQINGRQSPSONE-----RWAPKZHLIED--ETKQSEQRQSRNQST	1073
Db	492	ARGEKELUSPRTTPPAQKGAAPPAVGHGHELALSSPEDITQTQFEELKBERGULLAEQRD	551
QY	1074	T-----YFVYTESTDDKHLK-----FQP-----HFGQOECVSPYRSRGANGSETNR--	1114
Db	552	TGLGEKPLPADATEQCHPSAAIQVTPSGPVLGEHVEREKVVP--DSPGDKGS--TNRGP	609
QY	1115	-----VCSNHGINQNTVSQLCQEDDEDXDKPTNYRYERYEEBQ	1152
Db	610	DSGAVEKEKETWEERKOREABLGPENTAKARESEAEVKEDVIEKALEMEMETHPSDEE	669
QY	1153	HEBERPTNYSIKYNEKRHWDPIDYSLKYATDI-----PSSQKQSFSPKSSSSG	1203
Db	670	GEETKABSFY-----QKHTQBAKASPKSREALGGRDLFGQKAPKETATFSLSLAT	722
QY	1204	QSSKTEHMSSENSTPSPSNAKZ-----QONQH-----PSSAQSRSGQ-----PQ	1244
Db	723	PAGATEHVSYIQBTIPGYSETQIISDEIHEDPERPAPRPFSTYDLSQEPGPGPF	782
QY	1245	KAATCKVSSINQETIQYCYVEDTPICFRCSSLSLSSAADEICGNQTTQEAASANTLQI	1304
Db	783	EASQAADSAVPASSKTYGAPETELTYP--PNMVAAPLAEBE-----HVSSAITSITE	832
QY	1305	AETKGLGTRSAED--PVSEVPVAVSQHPRTKSSRLQGSLSLSSBSARHKAVE-----	1358
Db	833	CDKLSSEFATSVAEDQSVASLTAPQTEETGKSLLLDTVTGPSRTEATQGLDYVFSAGT	892
QY	1359	KPSKSGAQRTP--KSPPEHYVQETPLMFSTCTSVSSLDSFE-----SRSIASS-----	1404
Db	893	ISTPISLEEDKGFKSP-----CEDPSVTGESEKKGETVGRGLSEKAVGKE	939
QY	1405	-----VQSEPCGMVSGIISPSDLPDPSGQTPWPS-----RSKTPPP	1441
Db	940	EKYVVTSEKLSGGYAAVFG-----APGHTLPGEPALGEVEERCLSPDDSTVKMASPPP	993
QY	1442	--PPTAOTKREVPKNKAPTAEKE-----SPPKQA	1470
Db	994	SGPFSAAHT-----PFHOSPVDEKSEPRDFOEDSWGETKHSFGVSKEDSEQTVPKGPPEG	1049
QY	1471	-----AVNAAVQ-----RVQVLPDADTLHLHFATESPTDPGSCSSLSA	1508
Db	1050	TSEEGKPPTRPSQAQOMPYSIAGGQGTCTQLLPEQDKAIVFETG-----EAGSNLGA	1103
QY	1509	LSL-----DEPFIOKVELR-----IMPPVOENDNGNETSEOP	1542
Db	1104	GILPEGVRTSTEATEP--QNDVELRDTDOSLPEDAESLSVLVSPDPTTKQEAATRSP	1161
QY	1543	KESNENQEKAEKTIIDSEKOL-----LDDSDDDDDIILEECIISAMPTKSSRGKKPAQTAS	1599
Db	1162	CSLKEQQ-----PHKOLWPMVSPEDTQSLSFSEE-----SPSKET-----SLDISS	1202
QY	1600	KLPPPVV-----RKPSQLPVYKL-----LPSONRLQPKHVSFTFGDD	1637
Db	1203	KQLSPESLGTLOFELNLGKEERGFMKAEDDSCHLAPVPIPBPHRAIVSPSTDETPAGT	1262
QY	1638	MP-----RVYC-----VEGTPINFSTATSLDITIESSPNEIAA-----GE	1673
Db	1263	LPGGSFHSALSVDKRKHSPEBITGPGGHFWT---SDSSLTKSPESLSSPAMEDLAVWE	1318
QY	1674	GVRGGAQSGBEFKDXT-----IP-----TEGRSTDQAQGGKTSVTIPE	1712
Db	1319	GKAPGKEPELKETSETQOQQLPEKVAVVEQOLIITHQDKGALDENKEGQODKT-PE	1377

Qy	1713	LDDNKAEEGILLACINISAMPKXSHKPRVVKIMDVOQASASSAPNKNQLDGKKKEP	1777
Db	1378	QKGRDLDKXTAAELDKGPEPK	---
Qy	1773	TSPVKPTPQNTYRTRVRKNADSNNLNAERFVSDNKSQKQNLKNNKSDFNDKLPNNED	1832
Db	1429	LQQT---	QATEPRDRAQRD---
Qy	1933	RVRGSFAFDSPHYTIEGTPYCFSRNDSLSLDFDDDDVDSLREKAERLKAKE---	NKE 1889
Db	1473	R---	APESHIP---
Qy	1890	SEAKVTSHTELTNGOQSANKTOAIAKOPINRGOKPILQKQSTFFOSSKIDPRGAA---	1946
Db	1522	GET---	LDOEARTAEKQDTELKXDTQCGKSSFVEDKIT---
Qy	1947	TEKILQNFAIENTPVCFNSHSSLSLSDIQNNH-KENEPI---	KETEPDQCEP---
Db	1573	DSVEQDDGALEKTRALGLEESPAEGSKAREQEKYKKEQDVVQGWRETSP--	TRGEPVG 1630
Qy	2000	---	SKPQASGYAPKS--
Db	1631	GQKEPVPWAECKSPGEVRYWRDRDITLQDAYWRELSCDRKVWFPHLDDGQARFAYCE	1690
Qy	2051	KXPRL--	KGDNE---
Db	1691	ERESTFLDEGPDEQETPLQHTPRS---	PWTSDFKDFQBPLOKGLIVER---
Qy	2103	AIQEGANSIVSSLHQAAACLSQRASSDSLSLSKSGISLSPHLTPDOEEK---	P 2158
Db	1740	A---	---
Qy	2159	F---	TSNKGPRILKXPEKS---
Db	1759	FEIISPPASPMTGORVPS--	APGOESPVPDTESTAPMRNEPTTISWLAIEIPWPVKDR 1816
Qy	2187	---	---
Db	1817	PLPPAPLSPAPAPPTPAPEPHTVPFVSWGLAEYDSVVAVQEGAALEGXPSPLOKDYR	1876
Qy	2223	SISGRGTMIHIGVRNSSSSTSPVSKGPELKT---	PASKSPSE---
Db	1877	KARGEKEGCGAAPSPPSPKVPAGESLATRDTQEOTEPEOREPTTPPDERSFOYADI	1936
Qy	2264	---	---
Db	1937	YEQMMLTGLGPACTREPPPLGASGDVPHLSTKEEAGCNTSAEKETSPSPSQNLQSDT	1996
Qy	2309	P---	SPAQOPLSRFIQSPGRN---
Db	1997	PAPSYASLAGVAPPQOEP--	DGPNVRESITPPAVPPRAPLSKDLGLPPLNGSTVSCS 2054
Qy	2349	SSPSTASTSGSGKMSYSPGQMSQQMLTKOTGLSKNASSIPRSESASKGLNQ---	MN 2405
Db	2055	PDARTTSPKETGRHW---	DDOTNDSLEKGAREQEPKETRSPSPHPMPMG 2103
Qy	2406	NGGANKVLSRMSSTKSGCS---	---
Db	2104	HSLWPETEAYSSLSDDSHLGSVRLDPFASAFGSSLQAPPQLPSPAEPSPRSPAGSL	2163
Qy	2428	--	ESDRS---
Db	2164	AFSGDRALAVPGTPTRHDEYLEVTKAPSLDSSLIPOLPSPSPGGLLSNLPRASPA	2223
Qy	2475	---	---
Db	2224	LSEGSSEATTPISSVAERFPFGLEAAEQSAEGLSGKESAHSLSLWDTLP--	LSPA---
Qy	2518	NDGEPKAKHDIAKSHGESPSRLPINRSGTWKREHSHSSSLP---	RVTWRR 2566
Db	2279	---	---
Qy	2567	TGSSSSILSASSSESSEKASEDEKHYNS---	ISGTQSKENQVS---

Db 2319 TKPSPPLSPSGD--HEANGFGTSLNPPGFTVATAEKEBAEPHAWERGSWPEGHERSS 2376
 QY 2621 SPYN--STQTVSSGATNGAESKTLIYQM:-----APAVSKT 2654
 Db 2377 RPTDLSSEQLPRKSGSGGPPCSLSSEVAGFGCATDPRPHCGBELSPSFLNPPPLPST 2436
 QY 2655 EDVWVRLEDCEPI-----NPNRSGR-SPTGNTPTPVIVDSVSEKANKPNIKDSKDNQAKQNVNG 2709
 Db 2437 DSDSLSTEERLACKGRRRVGPGATGGCPWAD---ETPPTSASDSSGSSSDSDVPPE 2493
 QY 2710 SVMPTVGLNRLITS-----FIOVD-APDOKGT-EIKPGON-NPVPVSETNESP-----IV 2757
 Db 2494 TEPCPSITAEALDSDGDFLPVDRAGGVSGTHPRPGHDPPTLPDRPSPPRPDVC 2553
 QY 2758 ERTPFSSSSSS-----KHSSPSGTVAAVTTFN-----YNPGRKSSADSTS- 2799
 Db 2554 MADPEGLSSSGRVERLREKRGRRAPGRAPKASPARRLDIRKRSPTFGKGPVDRTSR 2613
 QY 2800 --ARPSQIPTPVANNKK 2815
 Db 2614 TVPRPRSTPSQVTSABEK 2631

RESULT 5
 SPEN DROME STANDARD; PRT: 5560 AA.
 AC Q8SX83: Q9NHN1; Q9NJ17; Q9U6C3; Q9VPL1; Q9VPL2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Split ends protein.
 GN SPEN OR CG18497.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RN SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), FUNCTION, ALTERNATIVE PROMOTER
 RP USAGE, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND MUTANTS B9 AND
 RP D57
 RC TISSUE=Embryo;
 RX MEDLINE=20025936; PubMed=10556082;
 RA Welllette E.B., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
 RA McGinnis W.;
 RT "spen encodes an RNP motif protein that interacts with Hox pathways
 RT to repress the development of head-like sclerites in the Drosophila
 RT trunk";
 RL Development 126:5373-5385(1999).
 [2]
 RN SEQUENCE FROM N.A. (ISOFORM 4).
 RP TISSUE=Embryo;
 RC MEDLINE=20157049; PubMed=10655223;
 RX Rebay I., Chen P., Hsiao P., Kolodziej P.A., Kuang B.H., Laverty T.,
 RA Suh C., Voas M., Williams A., Rubin G.M.;
 RA "A genetic screen for novel components of the Ras/mitogen-activated
 RT protein kinase signaling pathway that interact with the yan gene of
 RT Drosophila identifies split ends, a new RNA recognition motif-
 RT containing protein";
 RL Genetics 154:695-712(2000).
 [3]
 RN SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=20171275; PubMed=10704397;
 RA Kuang B.H., Wu S.C.-Y., Shin Y.-A., Luo L., Kolodziej P.A.;
 RT "split ends encodes large nuclear proteins that regulate neuronal
 RT cell fate and axon extension in the Drosophila embryo";
 RL Development 127:1517-1529(2000).
 [4]
 RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RP STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 [5]
 RN REVISIONS, AND ALTERNATIVE SPLICING.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kunkler J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.D., Whitfield E.J., Bayraktaroglu L., Serman B.F.,
 RA Bettencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 [6]
 RN SEQUENCE OF 424-2002 FROM N.A.
 RP STRAIN=Berkley; TISSUE=Embryo;
 RC MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Quarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celinker S.E.;
 RT "A Drosophila full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 [7]
 RN FUNCTION.
 RP MEDLINE=20253107; PubMed=10790398;
 RA Lane M.E., Elend M., Heidmann D., Herr A., Marzodko S., Herzog A.,
 RA Lehner C.F.;
 RT "A screen for modifiers of cyclin E function in Drosophila
 RT melanogaster identifies Cdk2 mutations, revealing the insignificance
 RT of putative phosphorylation sites in Cdk2";
 RL Genetics 155:233-244(2000).
 [8]
 RN FUNCTION ON EGF RECEPTOR PATHWAY.
 RP MEDLINE=20414403; PubMed=10959845;

Db	2509	EETHORISSDSAEHGGQDNQOGASTLADANRVQOQNLRLLCDGDDSSDEIRRNVMK	2566
Qy	793	QSYLGYGVFTNE--HDNRSDFNTGNTVLSPYLTNTVLPSSSSSRGSL-----	841
Db	2569	HSHFGRKSNSTRIADSESQSOPADLTIKQH---PIAPQEIQRQLSDEQKFKSR	2625
Qy	842	-DSRSSEKRSLEIRGI-----GLGNYPHATENPGTSSKGLQIS--	881
Db	2626	HDGNSSTEEKLTETRIKTELGDYFNSSBYTGTGLKBYSPETRKHKHKSRLXSST	2685
Qy	882	--TTAAQIAKMEEVSA---IHTSQB---DRESSGSTTELHCVTDERNALRRSAAHT	930
Db	2686	ADTSAAQTPLVMTPLTPIFDVHSSSECKTKFNFDDLKTECSSIPLEISAGER-----	2739
Qy	931	HSNTYNTKSENGNRUCSMFYAKLEYKRSDGNSGNGYKRGQMKPSTIESYSED	990
Db	2740	--SKHKERKEKREKLNRNMTATVPNSPTTNDTSSEKLSKE---ERHRLKSKKSKMDN	2794
Qy	991	ESFCSYGYQPADLAHKHS-----ANHMDNDGELDTPPINYSLKYSDEQJNSG	1039
Db	2795	S---CNTKIYNSGAHPSTSPSPATPSTAPTAQTSKRGDKWEIFGI--ISDESES--	2848
Qy	1040	RQSPQNERWARPKHIIEBIQ---SQORSRQNSTTYPVYTESTDDKHLKFPFHG	1094
Db	2849	-QPEQAE---TNKDIIPSSVSTGPIVSAALQYKQEPSTP--NSKNEEAHLQTVHEP	2902
Qy	1095	QBCVSPYRGRGANGETNRVGNHGINQVNSGLCOEDDYEDDKTNTVSRYSSEEQHE	1154
Db	2903	EQO-QOLLERSRLSGGSS---SSHADRER---HRREKREKREKSOREQQOIHQ	2951
Qy	1155	EEBR-----PTNYSIKYNEEKRVDOPI--DYSLK-----YATDI-----	1187
Db	2952	KSKVETKYDDDNSVDMEAGRALEAQLNSDFDTKPISEATSTAAATYRSDMTVFRS	3011
Qy	1188	PSSQKQSFPSKSSSQSSKTEHNSSESSENTPTSSNAKRONOLHPSSAQSRSGQOKAA	1247
Db	3012	DNEDNNSVDMTKQGVKSEQOECHKSKDKKKKKRSEKKEQKLL--QQORRESLPNVA	3068
Qy	1248	TKVLS-----SINQETIQYIC-----VEDTPICFRCSLSLSLSAEDEIG----	1288
Db	3069	TSGAPTPGKLTWNVQAASHADLOLDAKHISPPVC-KPSPLPCPLIGDDDDALHTPK	3127
Qy	1289	CNQTTOEADSANTLOIAIEKKG-----IGTRSAEDFVSEVPASVQHPRT	1332
Db	3128	AKPTPSSRGNDGLTPSREKPLISPIPKTPTIANSSTLTQSAETPVSSGTVIS----	3182
Qy	1333	KSRQLGSSLSSESARHKAVEFFSGAKSPKSGAOTPKSPEHYVQ-----ETPLMFSR	1386
Db	3183	-SALATPTPSSAAGVSA--PCLDNPSPTSAGQCKK--ESFIPGFGQLDDRISESA	3237
Qy	1387	CTSVSSLDSPEFSIASVQSEPCSGMVSGIISPSDLPSPGQTMPPSRKTPPPPPQTA	1446
Db	3238	VQISA--FENSTSLDNIADKIPVAS-----PPRATK--PLDKLE	3276
Qy	1447	QTRKREPKNAPTABKRESGPKQAAVNAVQVLPDADTLHFATESTP-----DGFS	1502
Db	3277	ESKSRV-----TISOBET--ESAVSALIGE-----SFGTSTTDSLQDMDE	3316
Qy	1503	SSLSALS-----LDBPFQKQVLRIMP-PVQEND	1532
Db	3317	MSVNELETFLVIAEPDEEALAAKAIETAGEPASLIEPEMEPEEAPDPDEAIE	3376
Qy	1533	NGNETSEQPKENEN-QEKEAKTID-----SEKLLDDSDDD-----TE	1573
Db	3377	SEPVVELDPEELNKAVQSLKHEDMDMDIKADTPQSERDLQIDTDTENPDEADSSGPSL	3436
Qy	1574	ILEECIIASMPKTS-SRKGKKPACTASKLPPPVARKP-----SOLPVYK--	1616
Db	3437	IDETVOSSSPEKISINNSPTPRETANIDIPNYESQPKLSESTPOPSVITKLPFLDTPK	3496
Qy	1617	-----LLPSQNRQLQPKHVGFTPGDDMPRVYCEGT-PINFSTATSLS-----DLT--	1661

QY 2533 SESPSRLPINSGTWKREHSHSSSL-----PRVSTWR-----TGSSS 2571
DB 4538 LPKGAQTPRSG--RNAQAKTDAVQIINAVGPRRSKORKTIGTQANTLIEBVTASNA 4595
QY 2572 SILGASSSESSKAKSE-----DKHVNISGT----- 2598
DB 4596 TVAASHLAPPGAGVESHVPQDRAKEVPSVSVVTPISTAPVSVAAFPVTFVPVAMPVKP 4655
QY 2599 ---KQSKENQVSAG--GTWRKIKENEF-----PTNSTQTVSSGATNG---AESKTLIYOM 2647
DB 4656 TMPQHPKKKALAAAEISYQAINISPSGUPMHTQTAAPATCKITGGVADAVSKALVDPV 4715
QY 2648 APAYS-----KTEDVWVRIEDCPINPRSRSTGNTPPVIDSVSKANPNIKSKDNQ 2701
DB 4716 TGVITAGMPOGKEGNLPAATAAPANS--SNEDGQAAPPPQLOHQOQOQHP--QOPPPQQ 4771
QY 2702 AKQNVGNSVPMRTVGLNRLTSF---IOVDAPQOKGTEIKPGONNPV-SETNESPIV 2757
DB 4772 ANLQINTLIFS---GLENPITALGKSVQLS-----TSAALINKPVSVLVKGNASQVI 4822
QY 2758 ERTPTSSSSSKHSPSGTVAARVTP--FNTNPSR-KSSADTSARPSQ-IPTPVNNNT 2813
DB 4823 QQ-----QOQP-IVAPAKQPIILQONPLPTVLHHAQHTVVRPPQPLKAHVLNRE 4870
QY 2814 KKRDSKTDSTESSGTQSPKR--HSGSYLVT 2841
DB 4871 KNIQOCLPTTKQAVAPQPPHAPSHGHMLLT 4900
RESULT 6
MINT_HUMAN
ID MINT_HUMAN STANDARD; PRT; 3664 AA.
AC Q9GT58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIA0929.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
RP RAR AND MTALL1.
RC TISSUE=Liver, and Pituitary;
RX MEDLINE=21231190; PubMed=1131609;
RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
RA Hon M., Evans R.M.;
RT "Sharp, an inducible cofactor that integrates nuclear receptor
RT repression and activation";
RL Genes Dev. 15:1140-1151(2001).
[2]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 294-3664 FROM N.A.
RA Rhodes S., Huckle E.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
RC TISSUE=Embryo, and Teratocarcinoma;
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Itoya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NEO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[5]

RP SEQUENCE OF 2002-3664 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hiroseawa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
[6]
RP INTERACTION WITH PP4D.
RX MEDLINE=21874127; PubMed=11867749;
RA Shi Y., Hon M., Evans R.M.;
RT "The peroxisome proliferator-activated receptor delta, an integrator
RT of transcriptional repression and nuclear receptor signaling";
RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
[7]
RP FUNCTION, AND INTERACTION WITH RBP5UH.
RX MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kostecka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway";
RL EMBO J. 21:5417-5426(2002).
[8]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
RX MEDLINE=22777836; PubMed=12897056;
RA Ariyoshi M., Schwabe J.W.R.;
RT "A conserved structural motif reveals the essential transcriptional
RT repression function of Spn proteins and their role in developmental
RT signaling";
RL Genes Dev. 17:1909-1920(2003).
CC -1- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative
CC regulator of the Notch pathway via its interaction with RBP5UH,
CC which prevents the association between NOTCH1 and RBP5UH, and
CC therefore suppresses the transactivation activity of Notch
CC signaling. Blocks the differentiation of precursor B cells into
CC marginal zone B cells. Probably represses transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins. May bind both to DNA and RNA.
CC -1- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
CC NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTALL1. Interacts with
CC RBP5UH; this interaction may prevent the interaction between
CC RBP5UH and NOTCH1. Interacts with the nuclear receptors RAR and
CC PP4D. Interacts with RAR in absence of ligand. Bind to the
CC steroid receptor RNA coactivator SRA.
CC -1- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -1- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
CC spleen and thymus. Expressed at intermediate level in kidney,
CC liver, mammary gland and skin.
CC -1- INDUCTION: By hormone 17-beta-estradiol (E2).
CC -1- DOMAIN: The RID domain mediates the interaction with nuclear
CC receptors (By similarity).
CC -1- DOMAIN: The SPOC domain, which mediates the interaction with
CC NCOR2, is essential for the repressive activity.
CC -1- SIMILARITY: Belongs to the Spn family.
CC -1- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -1- SIMILARITY: Contains 1 SPOC domain.
CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL; AF3556524; AAK52750.1; -


```
QY 1896 SHTELTSNOQANKTO---AIAKQPINRGQPKLQKQSTEPSQ---SKDIPDRGAATD 1948
Db 1514 DHKEEEOERQELFASRFLHSLIFEQDSKRLQHLERKEEDSDIFSGRIYKQOTSEGANSTT 1573
QY 1949 EKLQFAIENTPVCFSHNSLSLSDIDQENNNKENEPKIKETPPPOQSPSPKPOASGYA 2008
Db 1574 DSIQZ-----PVLPH-SRFMELTRMOQKEKEDQKPEVEKQEDTENHPKTPES---A 1623
QY 2009 PKSFHVE-DTPVCFSRNSSLSSLSIDSEDLLOECISSAMPKKKPSRLKGDNEKHSRPN 2067
Db 1624 PENKDSLTKPPPSVG-PFSVTVVTLSEAPSALE-----KTTGDKTVEAP-- 1666
QY 2068 MGGILGDLTLDLKDIOQRPDSEHGLSDSENFWDKAIQEGANSIVSLHQAAAACLSRQ 2127
Db 1667 ---LVTEKTEV-----FATVSEEAKPAPAPAVEQLEQVD---LPPG 1705
QY 2128 ASSDSDLSLSKSGISIGSPHLPTDDEKPFNTKGPRLKPKQEKSTLTKKIESKSG 2187
Db 1706 ADPPKEAM-MPAGVEEGSGDQPPYLDKAPPT-----PGASFQAESNVDPB-- 1752
QY 2188 IKGKKVVKSLITKVRNSBISQMKPLOANMPSISRGRTMIHPGVNRNSSTSPVS 2247
Db 1753 -----PAQKSEANE-----PKAEKPDATADAEPDANGKAEAPESQPPASEDLEVD 1808
QY 2248 KKGPLKTPAKSPSEGTATTSPRGAKPSVKSLSVARTSO----- 2291
Db 1761 K-----PAQKSEANE-----PKAEKPDATADAEPDANGKAEAPESQPPASEDLEVD 1808
QY 2292 --IGSSKAPRSRSDTSPRPA-----QPLSRPIQSPGRNISIPGRNGISPPNKLSQ 2344
Db 1809 PFVAAKDKPKNKS-KRSTKTPVQAASVIVKPKVTRKSERIDREKIKNSNPRGPAQKLE 1867
QY 2345 LPTSTSPSTKSGSGKMSYTPGROMQONTKTGTLKSNASSIPRSASKGLNQNM 2404
Db 1868 LKMAEKITRTASKNSAADLSEHPSPSLSR---TRRRNV-----RSVYATMGDHE- 1915
QY 2405 NNGNANKVKELSRMSSTKSGSDRSRERPVLRQSTFIKEAPSPTLRKLSSAS--- 2461
Db 1916 -NRSPVKPEVPQRYTRKRLERELQEAATAVPTPRRG-----RPPKTRRADEEENEAK 1969
QY 2462 --FELSPPSSRPAP-----TRSQATPVLPSPSL----- 2488
Db 1970 EPAETLKPEGHRSPRSQKTAAGGPGQKGNPKVDATREATTEV-GPQIGVKSSM 2028
QY 2489 -PDMSLSTHSSVQA-----GHRKRLPPNLSPTIEYNDGRPA--KHDIARSHSPSLP 2540
Db 2029 EFKAAEEEAAGSQKDRDKDAGTDKNPPETAP--VEVVEKKPAPEKNSKRGSRN-SLA 2086
QY 2541 INRSCTWKRHSKSSSLPRVSTWRTGTSSSILSASSESEKAKSED-----EKHVNIS 2596
Db 2087 VDKASLKNVDAVS---PRGAAQAGERESGVAVSPKESPOKEDGLSGLKSDPVD 2143
QY 2597 GTQSKENQVSAKGTWRKIKENEFPTNSTGTVSSGATNGABSKTLIQ-----MAP- 2649
Db 2144 PKPEKEKEDVSAGP--SPEATOLAKQMELEQAVEHIAKAEASASAAAYKADAPGLAPE 2201
QY 2650 -----AVSKTE-----DVWVRTEDCPINNRSRSGPTGNTTPV----- 2682
Db 2202 DRDKPAHQASETELAAAGSIINDISGEPENPAPPPYPGESQDLOQDPAGQAALQPSSE 2261
QY 2683 -----IDSVSEKAMPNIK-----DSKDNQAKQNVGSGVPMRTVGLN 2720
Db 2262 GNETDEAVSGILETEAATSSRPVNPAPDPSAGPTDTKEARNSSETSHSVPEAKGSKVE 2321
QY 2721 RUTSFQVADPOKGTIEKPGQNN-----FVP--VSEIN-----ESPIVE-----RT 2760
Db 2322 EVT-LVRKDKGRKTRSRKRNKTNKKVAVPESHVPSNQAGGSPAAANETTQVHPEA 2380
QY 2761 PFSSSSSSK-HSPSPGTVAAVTV--PFNVNPS-----PRKSADSTSRPQIPTP 2808
Db 2381 PQBEKQSEKXHPSTPPQSCSTSDLSKISTENSQEISVEERTTKASVPPDLPPPPQ-PAP 2439
```

```
QY 2809 VNNNTKRDSDKTDSTESSGTQSP 2831
Db 2440 VDEEPCARFVRHSIIESDPVTPP 2462

RESULT 7
MINT MOUSE
ID MINT MOUSE STANDARD; PRT: 3644 AA.
AC Q62504; Q80TN9; Q99PS4; Q9QZW2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mx2-interacting protein (SMAR1/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIAA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC TISSUE=Testis;
RX MEDLINE=99379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel mx2 binding protein, recognizes and
regulates the rat osteocalcin promoter.";
RL Biochemistry 38:10678-10690(1999).
RN [2]
SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
Ono M., Ohnato N., Saito R., Sakazume N., Sano H., Sasaki D.,
Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
Watahiki A., Muramatsu M., Hayashizaki Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura K., Kimura K., Kawaichi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
J/Su(H) activity.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 318-578 FROM N.A.
RC TISSUE=Cochlea;
RX MEDLINE=72737053; PubMed=9119401;
RA Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
Hamel C., Fzames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
Weil D., Puol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
unconventional type I myosins.";
RL Genomics 40:332-341(1997).
RN [5]
SEQUENCE OF 2598-3644 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=72579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
II. The complete nucleotide sequences of 400 mouse KIAA-homologous
cDNAs identified by screening of terminal sequences of cDNA clones
randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
RN [6]
TISSUE SPECIFICITY.
RX MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,
Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
```

FT	DOMAIN	1	574	DNA-BINDING.		
FT	DOMAIN	6	81	RNA-BINDING (RM) 1.		
FT	DOMAIN	336	416	RNA-BINDING (RM) 2.		
FT	DOMAIN	439	514	RNA-BINDING (RM) 3.		
FT	DOMAIN	518	590	RNA-BINDING (RM) 4.		
FT	DOMAIN	559	575	COILED COIL (POTENTIAL).		
FT	DOMAIN	822	850	COILED COIL (POTENTIAL).		
FT	DOMAIN	1185	1206	COILED COIL (POTENTIAL).		
FT	DOMAIN	1509	1544	COILED COIL (POTENTIAL).		
FT	DOMAIN	1607	1627	COILED COIL (POTENTIAL).		
FT	DOMAIN	2216	2704	RID.		
FT	DOMAIN	3478	3644	SFOC.		
FT	DOMAIN	2138	2462	INTERACTION WITH MSX2.		
FT	DOMAIN	2706	2845	INTERACTION WITH RBPSUH.		
FT	DOMAIN	125	277	ARG-RICH.		
FT	DOMAIN	236	326	SER-RICH.		
FT	DOMAIN	648	721	TVR-RICH.		
FT	DOMAIN	702	832	ARG-RICH.		
FT	DOMAIN	2101	2233	ALA-RICH.		
FT	DOMAIN	2377	2518	PRO-RICH.		
FT	DOMAIN	2950	3475	PRO-RICH.		
FT	VARSPLOC	618	640	Missing (in isoform 2).		
FT				/FTIC=VSP_008564.		
FT	VARIANT	348	348	I -> T.		
FT	VARIANT	762	762	S -> F.		
FT	VARIANT	773	773	S -> F.		
FT	VARIANT	933	933	S -> L.		
FT	CONFLICT	754	754	R -> G (IN REF. 3).		
FT	CONFLICT	1524	1524	D -> A (IN REF. 3).		
FT	CONFLICT	1560	1560	H -> Y (IN REF. 3).		
FT	CONFLICT	1570	1570	F -> L (IN REF. 3).		
FT	CONFLICT	1574	1574	R -> G (IN REF. 3).		
FT	CONFLICT	1609	1609	Q -> R (IN REF. 3).		
FT	CONFLICT	1659	1659	I -> V (IN REF. 3).		
FT	CONFLICT	1669	1669	S -> F (IN REF. 3).		
FT	CONFLICT	1705	1705	V -> A (IN REF. 3).		
FT	CONFLICT	1815	1815	A -> V (IN REF. 3).		
FT	CONFLICT	2097	2097	G -> A (IN REF. 3).		
FT	CONFLICT	2201	2202	MISSING (IN REF. 3).		
FT	CONFLICT	2322	2322	A -> V (IN REF. 3).		
FT	CONFLICT	2385	2385	P -> Q (IN REF. 3).		
FT	CONFLICT	2502	2502	R -> K (IN REF. 3).		
FT	CONFLICT	2505	2505	E -> K (IN REF. 3).		
FT	CONFLICT	2519	2519	T -> N (IN REF. 3).		
FT	CONFLICT	2554	2554	T -> S (IN REF. 3).		
FT	CONFLICT	2679	2698	LVSTPAGVN -> VGEHPWARD (IN REF. 3).		
FT	CONFLICT	3010	3010	L -> P (IN REF. 3 AND 5).		
FT	SEQUENCE	3644	AA; 398750 MW; 9CTEC49A8-A7DA4A	CRG64;		
Query Match						
Best Local Similarity 3.0%; Score 434.5; DB 1; Length 3644;						
Matches 524; Conservative 321; Mismatches 894; Indels 1069; Gaps 128;						
QY	727	RNLNMRPAK-YKDANIMSPGSSLP	SLHVRKQKALEALDAQHLSETPDNLSPK--A	783		
Db	226	REVRGRPERSYQHSRSRSPHSS	-----QSRNQSPQRLA	259		
QY	784	SHRSQRHKQSLGIDYVDFTNRHDDNR	DNDFNTGNMTVLSPLVNTTLPSSSSSRGSLDS	843		
Db	260	SQASRPTSPSGSG	-----SSRNTDS	298		
QY	844	SRSEKDRSLERBIGLGNYPATNPOTS	KRGQLGISTTAQIAKVMEE	893		
Db	299	SSTASDDSPAR	-----SVQAAVPAPTSQLLSLEKDEPRKSGIK	339		
QY	894	-----VSAIHTSQED	-----RSSGSTT--ELHCVTDERNAL	934		
Db	340	VQNLVPRSIDTSLKDLGPHFEKFKG	CKTSVQIHGASERYGLVFRQOEDKALTASKG	399		
QY	935	YNF	-----TKSN	-----SNRTCSM	949	
Db	400	KLFFGQIEVTAWGVPETESNEFPRL	DERIDEFHPKATRTLFIGNLEKITYYHDLNIF	459		

QY 950 -----PYAKLEY-----KSSNDLSKSVSNDGYGRGQWKP 981
Db 460 QRFGEIVDIDIKVNGVPOYAFLOYCDIASYCAIKAWDGYLGNRLKLGFGK---SMP 516
QY 982 S-----IESYBDDDESK-----FCYGOY-----PADLAHIIHSAN 1012
Db 517 TNCVMDGLSNVSDQYLTHFRYGVVVKVFORLGMALVYSEIEDAQAAYKETGR 576
QY 1013 HMDNDCGLD-----TPINYSLKYSDQL-----NSGRQSPSONERW---ARP 1052
Db 577 KIGNKIKVDPANRESQAFTHCEKSGQDWDYFEMLTERRAQMAQSHEDWSADAQS 636
QY 1053 KHIIEDEIKOS--EORQSR---NOST--TYPVYTESTDDKHLKQPHFGQOECVPSYRSR 1105
Db 637 PHKCREERRGVYSOERTYENVRTPCTYP-----EDSRDYPAR 677
QY 1106 GANGSETNRVSGHGNQVQSLOCDEDDYED---DKPTNYSE-----RYSEE 1150
Db 678 G-----REFYSEWETYQGEYDSRYDEPREYRSDPYEQDIREYSYR 722
QY 1151 ECHEEERPTNYIKYNEEKHVDOPIDYSLK-----YATDIPS--SOKQS 1194
Db 723 QREERERE-----RFSDRCHERPERSQSPVHLRPPQSPGVSPAHSERLPSDSERRL 777
QY 1195 PSFSKSSSGQSSK-----TEHMSSENSTSTP-----SS 1223
Db 778 YRSSERSGSCSVSPRYDKLEKARLERYTKNEKADKERTFDPVERERRIRVKEKGE 837
QY 1224 NAKRONQLHPSSAQSRGPOKAATCKVSSINOETIQYCVEDTPICFSRCSLSSLSA 1283
Db 838 KDKAERQKRGKAHSPSQP-----SETEQENDREOSP-----EXPRG 875
QY 1284 EDEICNOTTOBADSANTLO-----IAEIKGICTRSAEDPVSEVPVAVSHQHPRTKSSRL 1337
Db 876 STKLSRDRADKEGPAKRLLELVPCVVLTRVKEKGVIEHPPPEKRLKARLGRDITTKASAL 935
QY 1338 -----CGSSLSSESARHAYE---FPSGAKSPSKSG-----AQTP----- 1369
Db 936 DOKPOAAOEPKAPKDPARKALREKVLPSHAEVGEKERTKRLKHLKAEQIPELSALDLE 995
QY 1370 -----KSPPEHYVQETPLMFSR----- 1386
Db 996 KLEARKRPADSLKIEQKPEIKTSPB--TEDTRILLKQPDTRSDGVLLREGESERK 1053
QY 1387 -----CTSVSLDSFESRIASSVQSEPCSMVSGIISPDLPSP 1427
Db 1054 PYRKEILKRESKTKLERLNSALSPKQDPAVASGSGRPSDDVHAGL---GELTHGS 1110
QY 1428 GQTMPPSRKTPPPPTQATQKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPDAOT 1487
Db 1111 VETQTOPKKAIPSKQPKQL--OLLENQGP--EKEE-----VRKNYCRPREP 1155
QY 1488 LHFATESTPDGFCSSSLSALSLEPFTQKDVLRIMPVVOENDNGNETSEQPKESNE 1547
Db 1156 AEHRAGQEPHGNABEKLGI-----DID-----HTQYRKQMEQSRKQRM 1198
QY 1548 NOEKEAETIDSEKOLLDDSDDDIE---ILEECIISAMPTKSRKGGKPAQATASKLPPP 1604
Db 1199 MEIAKAEFGSGPKOV-----DDYERRSLVHEV-----GKPPQDVTDDSPS 1240
QY 1605 VARKPSQPVYKLLPSQNR-----LQPO--KHVSFTPGDD---MPRVYCV 1644
Db 1241 KKERTDHDV--FDICTKRENYRSSROI SEDSERTSCSPSVRHSFPHDDDDPRGSPRLVSV 1299
QY 1645 EGTG-----INFSTATSLDLTIESP--- 1665
Db 1300 KGSPPKDEKGLFYPNAAVRDDPLKCNPYDSGKREQTADTAKLVLNLSGEFSPSWDDPM 1359
QY 1666 -----PNELAAGEVGRGAQSGEFKRDITPTBGRS--TDE-----AQGGKTS-- 1707
Db 1360 KQDPSRDFVSFFNVIKRDLSR-----KRSVRDLEFGEVPSDDEDAEHRQSFRASSF 1413
QY 1708 VTIPELD-----DNKAERGDILAEBCINSAMPKGGKSHKPFVRVKIMDQ-----VOQASAS 1756

Db 1414 YDSPRLSFLLRDRDQKLREDD---ERLASSLERKNFYS--FALDKTITPTTKALLERAKSL 1469
QY 1757 SSAPNKN-----QLDGKKKKPTSPVKPIPONTEYRTRVRKNADSK----- 1796
Db 1470 SSSREENWGFLOWDFANFRNNKDKKVKDSAPRPIPSWYMKKKKIRTUSEGKLDKDKDB 1529
QY 1797 -----NNLNAERVFSDN---KDSKK--QNLKNNSKDFNDKLPNNEDVRGSAFADSPH 1844
Db 1530 RREEQERQELFASRFLHSSIPEODSKRIQHLERKSEE--SDPPPG---RLYGRQASEGAN 1585
QY 1845 HYTPISGTYPCF--SRNDSLSSLDFFDDDDVDLSREKAELRK-----AKENKES 1891
Db 1586 STSDSVQEPVVLPHSFHFMELTRMQQKEKQKQKPEAEKOESEPTHPKTPPEPAETKEPE 1645
QY 1892 AK-----VTSHTBLTSGQSA--NKTQAIKQAPINRG---QPKPILO----- 1928
Db 1646 PKAPVAGLPAVITITVTPPEPASSAPEKAEAEAAEASPAGERPAEPAPVSEETKLVSPEV 1705
QY 1929 -----KQSTFP--OSSKDIPDRGAATDEKQLONPAIENTVPCFSIN----- 1966
Db 1706 SVPEQPROSDVPPGDSRSDQSDAALAPSAPOESAATDAVPCVNAEPLTPGTTVSQVES 1765
QY 1967 -----SSLSLSDIDQENNNKENEPKETTEPPDS--QGEPSKPOASGVA-----PKSPH 2013
Db 1766 SVDPKPSSPOPLSKLTQRSEEAEGKV---EKDPTTSTEPDATONAGVASEAQPPASED 1822
QY 2014 VEDTP--VCFPSRNSLSSSIDSEDDLQECISSAM--PKKKPSPRLKGDNK--HSP 2065
Db 1823 VEANPPVAAKDRKTNKSKSKTS---VQAAAASVVEKPVTRKSERI--DREKLKRSSSP 1876
QY 2066 RNWGGILGE---DITLDLKDIOQP-----DSEH----- 2090
Db 1877 R-----GEAKLLELMKWEAEKILTRASKSGGDTTEHPSPSLPSRRRANRVSVATMT 1930
QY 2091 ---GLSPDSENFDWKAI----- 2104
Db 1931 DHESRPAKEPVQPRVTRKRLERLEQAVVPTTPRRGRPPKTRRAEEDGHERKEPA 1990
QY 2105 -----QEGANSIVSSLHOAAAAACLSE-----QASSDSLSLKSG--ISLGSFHL 2150
Db 1991 ETPRPAEGWRSRPSQSAASAAAPQCKRGRNEQVAAEAAGAAQASTREGNPKRGEEAA 2050
QY 2151 TPQDEEK--PFTSNKGP-----RIL--KPOEKSTLETKIESESKGKIGKKVYKSLITG 2201
Db 2051 SEPKDRRDSPTDKSGPDTTFVEVLERKPEK--TYKSKRGRANSTRSGMDRAAHQORLEM 2109
QY 2202 KVR-----SNSEISGQMKQPLQANMPSISRGRWMIHPGVRN-----SSST 2243
Db 2110 AARAAGQAADKEAGPAAASPOESSESPOKSGSS---POLANNPADPDREAEESESAST 2165
QY 2244 SP-----VSKKGPELKTPAKSPSEGTATTSPRGAKPSVKSELSPV 2285
Db 2166 APPEGTOLARQIELEQAVQNIKILPEPSAAAASKGTA---TATAASEEPAPHEGHKP- 2221
QY 2286 ARQTSQ-----IGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPG----- 2325
Db 2222 AHQASETELAAGIISIDASEPENFSAPPSVP--PGSQTHPREGMEPGLHEAESGILE 2279
QY 2326 ---RNSISPGRNGISPNKLSQLPRTSSPSTASTKSSGSGKMSYTSPPGRMSQOULTKO 2381
Db 2280 TGTATESAPQVSALDPPEGSADTKETRGNSGDSVQEAQSKAEVTPPRDKGKQKTRR 2339
QY 2382 TGLSKNASSIPRSEBSASKGLNQMNNGANKVKELSRMSSTKSSGSESRSRPERVLRQS 2441
Db 2340 -----RKNANKV--VAITETRASEAQTQSESPAASEAT 2373
QY 2442 TFIKEASPTLRRKLESASFESLPSRSRSPASPTRSOAQTPVLSPLPDLKSLTHSSV-- 2499
Db 2374 AATPEAQ-----EE-----KPSEKPPSPAECTFDPFKTP--PASELSQENSAAE 2417
QY 2500 ---QAGWRKLPNLSPTIEYNDGRPA--KRHDIAKSHSESP-----S 2537

Db 551 DTGLG-----DKPPLD-----TAEAGPPSTAIQOTPPSVPLGQEE 587
Qy 956 YKSSNDLSNVSSNDGCKGQKMPSTIESYSEDDSKFCYGYPADLAHITHSANMD 1015
Db 588 HVMKEKELVVPBEEQSGKRG-----LDSAEETEEK----- 620
Qy 1016 DNDGELDTPIYSLKYISDEQLNSRQSPSQNRWARPKHIIIEIDIKOS--EQRQSRNQST 1073
Db 621 -----DI-----WEEKQREARLPDRTE--AREES--EPEVKEDVIEKAELEEMEE 663
Qy 1074 TYPVYTESTDDKHLK--POPHGQOECVSPVRSRGANG-----SETNRVCSN- 1118
Db 664 VHPDEEBEDATKABGFOYKIMQPLKVP--RSREAFGREGLGQKAPKESKETSFLSSL 722
Qy 1119 --HGINQVNSQLCOEDDYDDKPTNYSE--RYSEBEOHEE--EERF-----TNYSI 1164
Db 723 TTPAGATEHVSY-----IQDEIIPGSETIEQIISDEIHDEEERAPPRFHTSIYDL 775
Qy 1165 KYNBEKR--HYDQPIDYS-----LYATDI--PSSOKQSPFSKSSSQ 1204
Db 776 PQEGAGFEASQFADSAVPATSGKVGTPETELTYPTNIVAAPLAEHEHVSSATSITEC 835
Qy 1205 SSKTEHMSSENSTPSSNAKRNQHLHPSSAQSRSGOPOKAATCKVSSINQETIQYCV 1264
Db 836 DKLSFATSVAEODQSALTA-----PQTEETGK--SSLLDITV----- 872
Qy 1265 EDTPICFSR-----CSSLSLSAEDBIG-----CNQTOBADSANTLQIAEI 1307
Db 873 --TSIPSKTEATQGLDVPSAGTISPTSLSEDKGPKSPCEDFSVTGESEKGEIIG- 929
Qy 1308 KKGITGRSAED-----PVSEVPASOHPRTKSSRLOGSSLSSESA 1347
Db 930 KGLSGERAVEEBEETANVEMSEKLCQYGTFTVPSAGHALHP-----GEPA 976
Qy 1348 RHKAVE--PPSGAKSPKSGAOTPKSPPEHYVQETPLMFESRCTSVSSLDSFESRSIASSVQ 1406
Db 977 LGEABERCLSPDDSVKASPPSPGPPS--ATHIPPHQSPVEEKSEBPQDFQAD----- 1028
Qy 1407 SEPCSGMGVGIISPDLPDSFGQTNPPRSKTPPPPPQ--TAQTKREVPPKNAATAKRE 1464
Db 1029 -----SMGDTKRTFGVGKEDAAEETVPGPEEGTLEKEEKVPPPRSPQAOEAP 1076
Qy 1465 SGPKQAAVNAVQRVQVLPDADTLHF-----ATESTPDGFC----- 1502
Db 1077 VNIDBGLTCT--IQLLPAQDKAIVFEIMEAGEPTGILGAHALPGGLRLTPOEPGKQ 1133
Qy 1503 -----SSLSALS-----LDPEFTIQDVELRIMPPV 1528
Db 1134 KDEVLRYPDORSLSPRDAESLSVSPSDTANOEPKSPGCLTEQYLHKD-----RWPEV 1189
Qy 1529 QENDNGNETESQPKESNENQEKAEKTIIDSEK-----DILDD 1566
Db 1190 SPEDTQSLSLS-----ESPSETSLDVSSKQUSPESLGTQLQFELNLGKEEMHLMQA 1243
Qy 1567 SDDD-----DIEILEECIISAMP-----TKSSRRKKGKPAQTASKLPPPVARK 1608
Db 1244 EDTSHHTAPMSVPEPHAATASPTDGTTRYSAQTDITDSDLRKSPPASSFSHTP--SGN 1301
Qy 1609 PSQLPVYKLLPQNRLQOKHVSFTPGDDMPRVYCVGTEPINF--STATSLDITIE----- 1663
Db 1302 GRYLPGAIPTSPDEHILTPDSSFESKSP--ESLPGP--ALEDIAIKWEDKVPGLKDRTEQKKE 1359
Qy 1664 -SPPNEL-----AAGEVGRGAQSG--EFEKRD 1698
Db 1360 PFPEKDEVLOQDKTLEHKEVVEPKDTAIVQKDEALHVKNVAVQODKALEQKGRDLEQD 1419
Qy 1689 T-----IPTGEGSTDBAOGGKTSVITIPELDNDKABEGDILAEICINSAM--PKGSKHX 1739
Db 1420 TALEQOKALEPKDKLEEKKALEQKDKIPEKDKALEQKDTALEQDKKALEPKDKDLE 1479
Qy 1740 -----PFRVKITMDQVQAASASSASAPNKNQLDGKKKKTSPVKPIPONTVEYT 1787
Db 1480 QKDRVLEQKEKTPBEKQKALDQ--KVRSEVHKAPEDTVAEKMDKORDLEQTKAPBQHQAOE 1533

Qy 1788 RVRKNADSKNNLNAERVFS--DNKDSKKON--LKONSKDFNDKLPNNEDRVRGSPAFD- 1841
Db 1539 QDKVSEKKQDALEQKYKQWALGQKDEALEQNIQALENHQTOQESLVQEDTKPKWLEB 1598
Qy 1842 -SPHYTPIEGTYPYCSRNDSSSLDDDDVDLSREKAEALRAKENKE----- 1889
Db 1599 KSEPKVAME-----EKLEALLEKTKALGLEBSLVQEGRARQEKEKYWRGQDVWOE 1649
Qy 1890 -SEAKYTSHTELTSNOOSANKTOAIKQPIN--RGQPKPILOKQSTPEQSS----- 1937
Db 1650 WQSTSTREBEPAGEQKELAPAWEDTSPQDNRYWRGREDVALEQDITYWELSCERKWWPF 1709
Qy 1938 KDIDPRGAA--TDEKLONFAIENTPVCFSHNSSLSLSDIDQENNNKENETIKE--TBP 1992
Db 1710 HELDGGQARPHYTEERESTF-----LDGGPDDEQEVPLREHATRS 1749
Qy 1993 PDSQG-----EPSPQASGYAPKSPHYE-----DTPVCFSRNSLSLSISDEDDLQ--E 2041
Db 1750 PWASDFKDFQESPO-----KGLEVERWLASPV-----GLPPEEDKLTSPPE 1794
Qy 2042 CIS-----SAMPKKKPKSRLKGDNEXHSPRNMGILGEDLTLD--LKD1-----QRPDSE 2089
Db 1795 IISPPASPPEMWQORVPSAPGOESPIPDPKLPHMKNEPTTTSWLADIIPWVPKDRPLPP 1854
Qy 2090 HGLS-----PDSN--FDW-----KAIQEGA-----NSIVS 2113
Db 1855 APLSPAPGPPTPAPESHTTAPFSGWTPEYDSVVAQVQEGAALEGGPYPLGSKDYKXVG 1914
Qy 2114 SLHQAAAACLSQASSDSLSLSKSGISLGSPPHLLTPDOER--KPFTENKG-----P 2165
Db 1915 EREBERAAPDK--SSHXKVPPEARKSHATTEPEQTEPEQREPTYPERSFOYADIYE 1972
Qy 2166 RILKPEKSTLETKKTJESKGIKGGKYKYSIITOKVRSNBSIISQMKQP-----LOAN 2220
Db 1973 QMWLTGLGPACPRE--PPLGAAGDWPPCLSTKEAAAGNTSABEKLSPISPKSLQSD 2029
Qy 2221 MPSIS-----RGRW--IHIPGVNSSSSTSPVSKGPKPLTPASKSPSGQATTTSPRGA 2274
Db 2030 TPTFSVAALAGTVPVPRXEPGPMESLTPPAPVPPAPILISKGPSPLNGNLIUSCPDRR 2089
Qy 2275 KPSVK-----SELSPVARQTSIQIGSSKAPSRSGSRDSTPSRPAQOPLSRP- 2320
Db 2090 SPFKESGRSHWDDSTSDSELEKGAQEAE-----KEAQSPSPHPHPGMSPT 2137
Qy 2321 -----TQSPGRNISPGNOISPP-----NKLSQLPRTSSPTASTKSSGSGKMSY 2366
Db 2138 LWPETEAHVSPPLXSHLGKARPSLDPFASAFGFSLEXAP--POXPSPAPRSPAPCGSLAF 2196
Qy 2367 T-----SPG-----RQMSQQLTKOTGLSKNASSIPRSES--ASKGLNOMNNGNANGKK 2413
Db 2197 SGDRALALAGPTRTTYDEYLEVTKAPSLD---SSLPOLPSPSSPGXPLLSN----- 2246
Qy 2414 VELSRMSTKSSGSDRSRERPVLRQSTFIKAPSTTLRKLUEESASFE----- 2463
Db 2247 --LPRPASPALSEGGSSSEATTPI--8SVAERFSPSLEAARQESGELDPGMEPAHXL 2300
Qy 2464 ---SLSPSSRSPASPTESQATPVLPSPLDMSLSTHSSVQAGGWRKLPNLSPTIET-- 2517
Db 2301 WDLTPUSPAP--PASLDLALAPAPSLPGDMGDGLPCHLECESEAAATEKPSFPQVPSDECAA 2359
Qy 2518 -----NDGRPAKRHDIAHSHSESPSRPINRSQW---KREHSKHSSSLPRVSTWR 2565
Db 2360 NGPTETSPNPPXPAP--AKAENEAEAAAXPAWERGAWPEGAERSRPRDTXLSPEQVPCP 2415
Qy 2566 RTGSSSILSASSE--SSEKAKSEBEKHVNSISGT-----KQSKENQVSA 2608
Db 2416 XGSGGGPPSSASPEVEAGPQGCXTEPRPHRKGELSPFLNPLPPLPSIDDDLDLSTEEVRLVG 2475
Qy 2609 KGTWRKI--KENEFSPNTNSTQTVSSGATNGAASKTIIYQMAPAVSKTEDVWVRIEDCP- 2665
Db 2476 RGGERRVGGFTTCXPFVTDTPPTISASDSGSSQS-----DSDVPPEETECCPS 2524

Qy 2666 -----INNPRSGRPTGNTPPVIDSVSKANPNIKMSKD 2699
 Db 2525 ITAEALDSDEGDFLPVXXGSGVTHPRPGHDP-----PPL-----PQDXRPS-- 2571
 Qy 2700 NQAKQNVGSPVMTVGLNRLTSTFIQVDAPQKGTETKPGQNNPVPVSETNESPIVER 2759
 Db 2572 -----PPRP-----DVCMADEPEGLSSESGEXERLXKXKVKQVRGVR 2608
 Qy 2760 TPSSSSSKHSPSGTVAARVTPFNPNPSPKSSADSTARSQIPTPNNNKKRDSK 2819
 Db 2609 AP-----GKDPVFXRLXLR-----GKSPFGSGXDRVSRXPSRSTQSVTPAEKXD 2661
 Qy 2820 TDSTESSG 2827
 Db 2662 GHSPMSXG 2669

RESULT 10
 PCLO_RAT STANDARD; PRT; 5085 AA.
 AC Q9JUS6; Q9JUL1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
 GN PCLO.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
 RX MEDLINE=20170257; PubMed=10707984;
 RA Fensler S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
 RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
 RT "Piccolo, a presynaptic zinc finger protein structurally related to
 RT bassoon.";
 RL Neuron 25:203-214 (2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fensler S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694.
 RX MEDLINE=21181819; PubMed=11285225;
 RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RT "An unusual C(2)-domain in the active-zone protein piccolo:
 RT implications for Ca(2+) regulation of neurotransmitter release.";
 RL EMBO J. 20:1605-1619 (2001).
 CC -I- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -I- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
 CC similarity).
 CC -I- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC Name=2;
 CC IsoId=Q9JUS6-1; Sequence=Displayed;
 CC IsoId=Q9JUS6-2; Sequence=VSP_003930, VSP_003931;
 CC -I- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -I- SIMILARITY: Contains a large conformational change.
 CC -I- SIMILARITY: Contains 1 C2 domains.
 CC -I- SIMILARITY: Contains 1 PDZ/DRH domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF138789; AAF07822.2; -
 DR EMBL; AF227534; AAF63196.1; -
 DR HSSP; P04410; 1A25.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding; IDA.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
 DR GO; GO:0005522; F:profilin binding; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR008899; Znf_piccolo.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF05715; Zfpiccolo; 2.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.
 DR PROSITE; PS0106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT TKPTN -> SKRRK (in isoform 2).
 FT /FTID=VSP_003930.
 FT Missing (in isoform 2).
 FT /FTID=VSP_003931.
 FT D-S-A: COMPLETE LOSS OF CALCIUM-BINDING
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
 FT BINDING ACTIVITY.
 FT D-S-A: COMPLETE LOSS OF CALCIUM-BINDING
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
 FT BINDING ACTIVITY.
 FT V-S-S: SMALL INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT M->S: INCREASED AFFINITY FOR CALCIUM.
 FT VW->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT QN->AA: MODERATE INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT A-S-S: NO EFFECT ON CALCIUM-BINDING
 FT ACTIVITY.
 SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
 Query Match 2.7%; Score 400; DB 1; Length 5085;
 Best Local Similarity 19.9%; Pred. No. 3.3e-07;
 Matches 387; Conservative 243; Mismatches 765; Indels 546; Gaps 83;
 Qy 1187 IPSSKQSPFSKSSSQSSKTEHMSSESTSTPSNAKRQNLHPS--SASRSQGPQ 1244
 Db 67 VPPAAASPSMHRKQELDSQAQPGKPPDPGRPTQPGLSKRTTDTFRSEKLPGRSP 136
 Qy 1245 KAATCKVSSINQETIQYCYVEDTPICFSCSSLSLSAE-----DETGCNQTQADS 1298
 Db 127 STSLKSKSKRTDFPKSKSMMPGFSDVNPUSAVSVVKNPFLLIDSDSASQETT 186
 Qy 1299 ANTQIAEIKGIGTRSAEDPV---SEVPASVQHPRTKSSRLQGS---SLSSSEARHAY 1352
 Db 187 KKQKVQKEQK-SEGMAKPPLOQPSKP-IPKQGVQVKEVIQDSDSPKSVSSQAQKVP 245

Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

(3)

SEQUENCE FROM N.A. (ISOFORM 1).

TISSUE=Brain stem; PubMed=8253844;

RA Chan W., Kordeli E., Bennett V.;

RA "40-KD ankyrin: structure of the major developmentally regulated

RT domain and selective localization in unmyelinated axons";

RL J. Cell Biol. 123:1463-1473(1993).

[4]

SEQUENCE OF 463-495 FROM N.A.

RA MEDLINE=9200921; PubMed=1833308;

RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,

RA Lux S.E., Ward D.C., Forget B.G.;

RT "Isolation and chromosomal localization of a novel nonerythroid

RT ankyrin gene";

RL Genomics 10:858-866(1991).

CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal

CC elements. Also bind to cytoskeletal proteins.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Name=1;

CC IsoId=Q01484-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q01484-2; Sequence=VSP_000267, VSP_000268;

CC Name=3;

CC IsoId=Q01484-3; Sequence=VSP_000268;

CC -!- TISSUE SPECIFICITY: Plasma membrane of neurons as well as glial

CC cells throughout the brain.

CC -!- PM: Phosphorylated at multiple sites by different protein kinases

CC and each phosphorylation event regulates the protein's structure

CC and function (Potential).

CC -!- SIMILARITY: Contains 23 ANK repeats.

CC -!- SIMILARITY: Contains 1 death domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

EMBL; X56957; CAA40278.1; -

DR EMBL; X56958; CAA40279.2; -

DR EMBL; Z26634; CAB42644.1; -

DR EMBL; M37123; AAA62828.1; -

DR PIR; S37431; S37431.

DR HSP; P42771; IDC2.

DR Genew; HGNC:493; ANK2.

DR MIM; 106410; -

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000488; Death.

DR InterPro; IPR000906; ZU5.

DR Pfam; PF00023; ank; 24.

DR Pfam; PF00791; death; 1.

DR Pfam; PF00791; ZU5; 1.

DR PRINTS; PR01415; ANKYRIN.

DR SMART; SM00248; ANK; 22.

DR SMART; SM00005; DEATH; 1.

DR SMART; SM00218; ZU5; 1.

DR PROSITE; PS50088; ANK_REPEAT; 20.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS50017; DEATH_DOMAIN; 1.

DR Cytoskeleton; Alternative splicing; Repeat; ANK repeat;

KW Phosphorylation.

FT REPEAT 63 92 ANK 1.

FT REPEAT 96 125 ANK 2.

FT REPEAT 129 158 ANK 3.

FT REPEAT 162 191 ANK 4.

FT REPEAT 193 220 ANK 5.

FT REPEAT 232 261 ANK 6.

FT REPEAT 265 294 ANK 7.

298 327 REPEAT ANK 8.

331 360 REPEAT ANK 9.

364 393 REPEAT ANK 10.

397 426 REPEAT ANK 11.

430 459 REPEAT ANK 12.

463 492 REPEAT ANK 13.

496 525 REPEAT ANK 14.

529 558 REPEAT ANK 15.

562 591 REPEAT ANK 16.

595 624 REPEAT ANK 17.

628 657 REPEAT ANK 18.

661 690 REPEAT ANK 19.

694 723 REPEAT ANK 20.

727 756 REPEAT ANK 21.

760 789 REPEAT ANK 22.

793 822 REPEAT ANK 23.

825 854 REPEAT-RICH REGION.

857 886 REPEAT A.

889 918 REPEAT A.

921 950 REPEAT A.

953 982 REPEAT A.

985 1014 REPEAT A.

1017 1046 REPEAT A.

1049 1078 REPEAT A.

1081 1110 REPEAT A.

1113 1142 REPEAT A.

1145 1174 REPEAT A.

1177 1206 REPEAT A.

1209 1238 REPEAT A.

1241 1270 REPEAT A.

1273 1302 REPEAT A.

1305 1334 REPEAT A.

1337 1366 REPEAT A.

1369 1398 REPEAT A.

1401 1430 REPEAT A.

1433 1462 REPEAT A.

1465 1494 REPEAT A.

1497 1526 REPEAT A.

1529 1558 REPEAT A.

1561 1590 REPEAT A.

1593 1622 REPEAT A.

1625 1654 REPEAT A.

1657 1686 REPEAT A.

1689 1718 REPEAT A.

1721 1750 REPEAT A.

1753 1782 REPEAT A.

1785 1814 REPEAT A.

1817 1846 REPEAT A.

1849 1878 REPEAT A.

1881 1910 REPEAT A.

1913 1942 REPEAT A.

1945 1974 REPEAT A.

1977 2006 REPEAT A.

2009 2038 REPEAT A.

2041 2070 REPEAT A.

2073 2102 REPEAT A.

2105 2134 REPEAT A.

2137 2166 REPEAT A.

2169 2198 REPEAT A.

2201 2230 REPEAT A.

2233 2262 REPEAT A.

2265 2294 REPEAT A.

2297 2326 REPEAT A.

2329 2358 REPEAT A.

2361 2390 REPEAT A.

2393 2422 REPEAT A.

2425 2454 REPEAT A.

2457 2486 REPEAT A.

2489 2518 REPEAT A.

2521 2550 REPEAT A.

2553 2582 REPEAT A.

2585 2614 REPEAT A.

2617 2646 REPEAT A.

2649 2678 REPEAT A.

2681 2710 REPEAT A.

2713 2742 REPEAT A.

2745 2774 REPEAT A.

2777 2806 REPEAT A.

2809 2838 REPEAT A.

2841 2870 REPEAT A.

2873 2902 REPEAT A.

2905 2934 REPEAT A.

2937 2966 REPEAT A.

2969 2998 REPEAT A.

3001 3030 REPEAT A.

3033 3062 REPEAT A.

3065 3094 REPEAT A.

3097 3126 REPEAT A.

3129 3158 REPEAT A.

3161 3190 REPEAT A.

3193 3222 REPEAT A.

3225 3254 REPEAT A.

3257 3286 REPEAT A.

3289 3318 REPEAT A.

3321 3350 REPEAT A.

3353 3382 REPEAT A.

3385 3414 REPEAT A.

3417 3446 REPEAT A.

3449 3478 REPEAT A.

3481 3510 REPEAT A.

3513 3542 REPEAT A.

3545 3574 REPEAT A.

3577 3606 REPEAT A.

3609 3638 REPEAT A.

3641 3670 REPEAT A.

3673 3702 REPEAT A.

3705 3734 REPEAT A.

3737 3766 REPEAT A.

3769 3798 REPEAT A.

3801 3830 REPEAT A.

3833 3862 REPEAT A.

3865 3894 REPEAT A.

3897 3926 REPEAT A.

3929 3958 REPEAT A.

3961 3990 REPEAT A.

3993 4022 REPEAT A.

4025 4054 REPEAT A.

4057 4086 REPEAT A.

4089 4118 REPEAT A.

4121 4150 REPEAT A.

4153 4182 REPEAT A.

4185 4214 REPEAT A.

4217 4246 REPEAT A.

4249 4278 REPEAT A.

4281 4310 REPEAT A.

4313 4342 REPEAT A.

4345 4374 REPEAT A.

4377 4406 REPEAT A.

4409 4438 REPEAT A.

4441 4470 REPEAT A.

4473 4502 REPEAT A.

4505 4534 REPEAT A.

4537 4566 REPEAT A.

4569 4598 REPEAT A.

4601 4630 REPEAT A.

4633 4662 REPEAT A.

4665 4694 REPEAT A.

4697 4726 REPEAT A.

4729 4758 REPEAT A.

4761 4790 REPEAT A.

4793 4822 REPEAT A.

4825 4854 REPEAT A.

4857 4886 REPEAT A.

4889 4918 REPEAT A.

4921 4950 REPEAT A.

4953 4982 REPEAT A.

4985 5014 REPEAT A.

5017 5046 REPEAT A.

5049 5078 REPEAT A.

5081 5110 REPEAT A.

5113 5142 REPEAT A.

5145 5174 REPEAT A.

5177 5206 REPEAT A.

5209 5238 REPEAT A.

5241 5270 REPEAT A.

5273 5302 REPEAT A.

5305 5334 REPEAT A.

5337 5366 REPEAT A.

5369 5398 REPEAT A.

5401 5430 REPEAT A.

5433 5462 REPEAT A.

5465 5494 REPEAT A.

5497 5526 REPEAT A.

5529 5558 REPEAT A.

5561 5590 REPEAT A.

5593 5622 REPEAT A.

5625 5654 REPEAT A.

5657 5686 REPEAT A.

5689 5718 REPEAT A.

5721 5750 REPEAT A.

5753 5782 REPEAT A.

5785 5814 REPEAT A.

5817 5846 REPEAT A.

5849 5878 REPEAT A.

5881 5910 REPEAT A.

5913 5942 REPEAT A.

5945 5974 REPEAT A.

5977 6006 REPEAT A.

6009 6038 REPEAT A.

6041 6070 REPEAT A.

6073 6102 REPEAT A.

6105 6134 REPEAT A.

6137 6166 REPEAT A.

6169 6198 REPEAT A.

6201 6230 REPEAT A.

6233 6262 REPEAT A.

6265 6294 REPEAT A.

6297 6326 REPEAT A.

6329 6358 REPEAT A.

6361 6390 REPEAT A.

6393 6422 REPEAT A.

6425 6454 REPEAT A.

6457 6486 REPEAT A.

6489 6518 REPEAT A.

6521 6550 REPEAT A.

6553 6582 REPEAT A.

6585 6614 REPEAT A.

6617 6646 REPEAT A.

6649 6678 REPEAT A.

6681 6710 REPEAT A.

6713 6742 REPEAT A.

6745 6774 REPEAT A.

6777 6806 REPEAT A.

6809 6838 REPEAT A.

6841 6870 REPEAT A.

6873 6902 REPEAT A.

6905 6934 REPEAT A.

6937 6966 REPEAT A.

6969 6998 REPEAT A.

7001 7030 REPEAT A.

7033 7062 REPEAT A.

7065 7094 REPEAT A.

7097 7126 REPEAT A.

7129 7158 REPEAT A.

7161 7190 REPEAT A.

7193 7222 REPEAT A.

7225 7254 REPEAT A.

7257 7286 REPEAT A.

7289 7318 REPEAT A.

7321 7350 REPEAT A.

7353 7382 REPEAT A.

7385 7414 REPEAT A.

7417 7446 REPEAT A.

7449 7478 REPEAT A.

7481 7510 REPEAT A.

7513 7542 REPEAT A.

7545 7574 REPEAT A.

7577 7606 REPEAT A.

7609 7638 REPEAT A.

7641 7670 REPEAT A.

7673 7702 REPEAT A.

7705 7734 REPEAT A.

7737 7766 REPEAT A.

7769 7798 REPEAT A.

7801 7830 REPEAT A.

7833 7862 REPEAT A.

7865 7894 REPEAT A.

7897 7926 REPEAT A.

7929 7958 REPEAT A.

7961 7990 REPEAT A.

7993 8022 REPEAT A.

8025 8054 REPEAT A.

8057 8086 REPEAT A.

8089 8118 REPEAT A.

8121 8150 REPEAT A.

8153 8182 REPEAT A.

8185 8214 REPEAT A.

8217 8246 REPEAT A.

8249 8278 REPEAT A.

8281 8310 REPEAT A.

8313 8342 REPEAT A.

8345 8374 REPEAT A.

8377 8406 REPEAT A.

8409 8438 REPEAT A.

8441 8470 REPEAT A.

8473 8502 REPEAT A.

8505 8534 REPEAT A.

8537 8566 REPEAT A.

8569 8598 REPEAT A.

8601 8630 REPEAT A.

8633 8662 REPEAT A.

8665 8694 REPEAT A.

8697 8726 REPEAT A.

8729 8758 REPEAT A.

8761 8790 REPEAT A.

8793 8822 REPEAT A.

8825 8854 REPEAT A.

8857 8886 REPEAT A.

8889 8918 REPEAT A.

8921 8950 REPEAT A.

8953 8982 REPEAT A.

8985 9014 REPEAT A.

9017 9046 REPEAT A.

9049 9078 REPEAT A.

9081 9110 REPEAT A.

9113 9142 REPEAT A.

9145 9174 REPEAT A.

9177 9206 REPEAT A.

9209 9238 REPEAT A.

9241 9270 REPEAT A.

9273 9302 REPEAT A.

9305 9334 REPEAT A.

9337 9366 REPEAT A.

9369 9398 REPEAT A.

9401 9430 REPEAT A.

9433 9462 REPEAT A.

9465 9494 REPEAT A.

9497 9526 REPEAT A.

9529 9558 REPEAT A.

9561 9590 REPEAT A.

9593 9622 REPEAT A.

9625 9654 REPEAT A.

9657 9686 REPEAT A.

9689 9718 REPEAT A.

9721 9750 REPEAT A.

9753 9782 REPEAT A.

9785 9814 REPEAT A.

9817 9846 REPEAT A.

9849 9878 REPEAT A.

9881 9910 REPEAT A.

9913 9942 REPEAT A.

9945 9974 REPEAT A.

9977 10006 REPEAT A.

10009 10038 REPEAT A.

10041 10070 REPEAT A.

10073 10102 REPEAT A.

10105 10134 REPEAT A.

10137 10166 REPEAT A.

10169 10198 REPEAT A.

10201 10230 REPEAT A.

10233 10262 REPEAT A.

10265 10294 REPEAT A.

10297 10326 REPEAT A.

10329 10358 REPEAT A.

10361 10390 REPEAT A.

10393 10422 REPEAT A.

10425 10454 REPEAT A.

10457 10486 REPEAT A.

10489 10518 REPEAT A.

10521 10550 REPEAT A.

10553 10582 REPEAT A.

10585 10614 REPEAT A.

10617 10646 REPEAT A.

10649 10678 REPEAT A.

10681 10710 REPEAT A.

10713 10742 REPEAT A.

10745 10774 REPEAT A.

10777 10806 REPEAT A.

10809 10838 REPEAT A.

10841 10870 REPEAT A.

10873 10902 REPEAT A.

10905 10934 REPEAT A.

10937 10966 REPEAT A.

10969 10998 REPEAT A.

11001 11030 REPEAT A.

11033 11062 REPEAT A.

11065 11094 REPEAT A.

11097 11126 REPEAT A.

11129 11158 REPEAT A.

11161 11190 REPEAT A.

11193 11222 REPEAT A.

11225 11254 REPEAT A.

11257 11286 REPEAT A.

11289 11318 REPEAT A.

11321 11350 REPEAT A.

11353 11382 REPEAT A.

11385 11414 REPEAT A.

11417 11446 REPEAT A.

11449 11478 REPEAT A.

11481 11510 REPEAT A.

11513 11542 REPEAT A.

11545 11574 REPEAT A.

11577 11606 REPEAT A.

11609 11638 REPEAT A.

11641 11670 REPEAT A.

11673 11702 REPEAT A.

11705 11734 REPEAT A.

11737 11766 REPEAT A.

11769 11798 REPEAT A.

11801 11830 REPEAT A.

11833 11862 REPEAT A.

11865 11894 REPEAT A.

11897 11926 REPEAT A.

11929 11958 REPEAT A.

11961 11990 REPEAT A.

11993 12022 REPEAT A.

12025 12054 REPEAT A.

12057 12086 REPEAT A.

12089 12118 REPEAT A.

12121 12150 REPEAT A.

12153 12182 REPEAT A.

12185 12214 REPEAT A.

12217 12246 REPEAT A.

12249 12278 REPEAT A.

12281 12310 REPEAT A.

12313 12342 REPEAT A.

12345 12374 REPEAT A.

12377 12406 REPEAT A.

12409 12438 REPEAT A.

12441 12470 REPEAT A.

12473 12502 REPEAT A.

12505 12534 REPEAT A.

12537 12566 REPEAT A.

12569 12598 REPEAT A.

12601 12630 REPEAT A.

12633 12662 REPEAT A.

12665 12694 REPEAT A.

12697 12726 REPEAT A.

12729 12758 REPEAT A.

12761 12790 REPEAT A.

12793 12822 REPEAT A.

12825 12854 REPEAT A.

12857 12886 REPEAT A.

12889 12918 REPEAT A.

12921 12950 REPEAT A.

12953 12982 REPEAT A.

12985 13014 REPEAT A.

13017 13046 REPEAT A.

13049 13078 REPEAT A.

13081 13110 REPEAT A.

13113 13142 REPEAT A.

13145 13174 REPEAT A.

13177 13206 REPEAT A.

13209 13238 REPEAT A.

13241 13270 REPEAT A.

13273 13302 REPEAT A.

13305 13334 REPEAT A.

13337 13366 REPEAT A.

13369 13398 REPEAT A.

13401 13430 REPEAT A.

13433 13462 REPEAT A.

13465 13494 REPEAT A.

13497 13526 REPEAT A.

13529 13558 REPEAT A.

13561 13590 REPEAT A.

13593 13622 REPEAT A.

13625 13654 REPEAT A.

13657 13686 REPEAT A.

13689 13718 REPEAT A.

13721 13750 REPEAT A.

13753 13782 REPEAT A.

13785 13814 REPEAT A.

13817 13846 REPEAT A.

13849 13878 REPEAT A.

13881 13910 REPEAT A.

13913 13942 REPEAT A.

13945 13974 REPEAT A.

13977 14006 REPEAT A.

14009 14038 REPEAT A.

14041 14070 REPEAT A.

14073 14102 REPEAT A.

14105 14134 REPEAT A.

14137 14166 REPEAT A.

14169 14198 REPEAT A.

14201 14230 REPEAT A.

14233 14262 REPEAT A.

14265 14294 REPEAT A.

14297 14326 REPEAT A.

14329 14358 REPEAT A.

14361 14390 REPEAT A.

14393 14422 REPEAT A.

14425 14454 REPEAT A.

14457 14486 REPEAT A.

14489 14518 REPEAT A.

14521 14550 REPEAT A.

14553 14582 REPEAT A.

14585 14614 REPEAT A.

14617 14646 REPEAT A.

14649 14678 REPEAT A.

14681 14710 REPEAT A.

14713 14742 REPEAT A.

14745 14774 REPEAT A.

14777 14806 REPEAT A.

14809 14838 REPEAT A.

14841 14870 REPEAT A.

14873 14902 REPEAT A.

14905 14934 REPEAT A.

14937 14966 REPEAT A.

14969 14998 REPEAT A.

15001 15030 REPEAT A.

15033 15062 REPEAT A.

15065 15094 REPEAT A.

15097 15126 REPEAT A.

15129 15158 REPEAT A.

15161 15190 REPEAT A.

15193 15222 REPEAT A.

15225 15254 REPEAT A.

15257 15286 REPEAT A.

15289 15318 REPEAT A.

15321 15350 REPEAT A.

15353 15382 REPEAT A.

15385 15414 REPEAT A.

15417 15446 REPEAT A.

15449 15478 REPEAT A.

15481 15510 REPEAT A.

15513 15542 REPEAT A.

15545 15574 REPEAT A.

15577 15606 REPEAT A.

15609 15638 REPEAT A.

15641 15670 REPEAT A.

15673 15702 REPEAT A.

15705 15734 REPEAT A.

15737 15766 REPEAT A.

15769 15798 REPEAT A.

15801 15830 REPEAT A.

15833 15862 REPEAT A.

15865 15894 REPEAT A.

15897 15926 REPEAT A.

15929 15958 REPEAT A.

15961 15990 REPEAT A.

15993 16022 REPEAT A.

16025 16054 REPEAT A.

16057 16086 REPEAT A.

16089 16118 REPEAT A.

16121 16150 REPEAT A.

16153 16182 REPEAT A.

16185 16214 REPEAT A.

16217 16246 REPEAT A.

16249 16278 REPEAT A.

16281 16310 REPEAT A.

16313 16342 REPEAT A.

16345 16374 REPEAT A.

16377 16406 REPEAT A.

16409 16438 REPEAT A.

16441 16470 REPEAT A.

16473 16502 REPEAT A.

16505 16534 REPEAT A.

16537 16566 REPEAT A.

16569 16598 REPEAT A.

16601 16630 REPEAT A.

16633 16662 REPEAT A.

16665 16694 REPEAT A.

16697 16726 REPEAT A.

16729 16758 REPEAT A.

16761 16790 REPEAT A.

16793 16822 REPEAT A.

16825 16854 REPEAT A.

16857 16886 REPEAT A.

16889 16918 REPEAT A.

16921 16950 REPEAT A.

16953 16982 REPEAT A.

16985 17014 REPEAT A.

17017 17046 REPEAT A.

17049 17078 REPEAT A.

17081 17110 REPEAT A.

17113 17142 REPEAT A.

17145 17174 REPEAT A.

17177 17206 REPEAT A.

17209 17238 REPEAT A.

17241 17270 REPEAT A.

17273 17302 REPEAT A.

17305 17334 REPEAT A.

17337 17366 REPEAT A.

17369 17398 REPEAT A.

17401 17430 REPEAT A.

17433 17462 REPEAT A.

17465 17494 REPEAT A.

17497 17526 REPEAT A.

17529 17558 REPEAT A.

17561 17590 REPEAT A.

17593 17622 REPEAT A.

17625 17654 REPEAT A.

17657 17686 REPEAT A.

17689 17718 REPEAT A.

17721 17750 REPEAT A.

17753 17782 REPEAT A.

17785 17814 REPEAT A.

17817 17846 REPEAT A.

17849 17878 REPEAT A.

17881 17910 REPEAT A.

17913 17942 REPEAT A.

17945 17974 REPEAT A.

17977 18006 REPEAT A.

18009 18038 REPEAT A.

18041 18070 REPEAT A.

18073 18102 REPEAT A.

18105 18134 REPEAT A.

18137 18166 REPEAT A.

18169 18198 REPEAT A.

18201 18230 REPEAT A.

18233 18262 REPEAT A.

18265 18294 REPEAT A.

18297 18326 RE

J. Cell Biol. 147:151-162 (1999).
 [2]
 RN SEQUENCE OF 552-4404 FROM N.A.
 RA Kraemer J., Wollam C., Wohldmann P., McGrane B.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
 RP TISSUE=Brain;
 RC MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:31-39 (1998).
 [4]
 RN SEQUENCE OF 4405-4439 FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dapkin L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gharatne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [5]
 RN SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Elliott G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Y19188; CAB60727.1; -;
 DR EMBL; AC004903; AAD20936.1; -;
 DR EMBL; AC004886; AAD21789.1; -;
 DR EMBL; AB011131; BAA25485.1; -;
 DR EMBL; BC001304; AAH01304.1; -;
 DR EMBL; AC004082; AAB97937.1; -;
 DR FIR; T06634; T06634.
 DR HSP; P04410; I425.
 DR GENE; HGNC:13406; PCLO.
 DR MIM; 604918; -;
 DR GO; GO:0005856; C:cytoskeleton; NAS.
 DR GO; GO:0045202; C:synaptic junction; ISS.
 DR GO; GO:0005509; F:calcium ion binding; ISS.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
 DR GO; GO:0005522; F:profilin binding; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001565; Synaptotagmin.
 DR PRINTS; PR00360; C2DOMAIN.
 DR PRINTS; PR00399; SYNAPTOTAGMN.
 DR SMART; SMO0239; C2; 2.
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT NON_TER 1
 FT DOMAIN 400 465
 FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT S -> SGNGLGIRVGGKEIFGHSGEIGAYIAKILPGSSAE
 FT QTGLMEG (in isoform 2).
 FT /FTid=VSP_003923.
 FT K -> KPTDGTGVSHPTGEIQ (in isoform 2).
 FT /FTid=VSP_003924.
 FT G -> GQVMVQNAS (in isoform 2).
 FT /FTid=VSP_003925.
 FT TAKKS -> SKREK (in isoform 2).
 FT /FTid=VSP_003926.
 FT Missing (in isoform 2).
 FT /FTid=VSP_003927.
 FT SEQUENCE 5147 AA; 563537 MW; CDS084990498CD3C CRC64;
 Query Match 2.7%; Score 387.5; DB 1; Length 5147;
 Best Local Similarity 19.3%; Pred. No. 9.6e-07;
 Matches 506; Conservative 313; Mismatches 961; Indels 837; Gaps 120;
 QY 718 IAMGSAALNLANRPAKYKDANIMSPGSSL-----PSLHVTKOKALEALD 765
 DB 1 IPAGNEADLSQ---SEERQIAAVMSRAQGLPKGVFPAAESPMSH-RKQ-----ELD 52
 QY 766 AQHLSSETFDNLSPKASHRSKQHKQSLYGVDFDTRHDDNRSDNFNTGNMTVLSPY 825
 DB 53 SSH-----PPKQSGRPP-----DPGRPAQGLSKSRITDTF 83
 QY 826 LNTVLPSSSSSRGSLDSSRSSEKDRSLRERGLGNYPATENPG-TSSKRGLOISTTA 884
 DB 84 RSEQLPGRSPFTISLKEKSRTD-----LKEHKSSMPPGLSEVNALSASVV 133
 QY 885 AQIAKVMEEYSIAHTSOEDRSSGSGTTELHCVTDE----- 918
 DB 134 VKRNPFDLSDSEASQEE-----TTKKQKVQEGKPEGIKPPLQQQPKPKQOG 188
 QY 919 --RNALRRSAAHTSNNTYNTKSENSTCSNP-----YAKLYKRS 959
 DB 189 PGRDPLQDQDTPKXIS-----SQQPEKIKSPGTGKPIQGTPTPTDTHAKLPQR- 240

Db 2165 ASSIESVVRKPEGVADTVSTDLISEKDVKKAKKETGNGII-----LE 2210
 Qy 2728 V-----DAPKOKGTEIKFGQNNPV---PVSETNESPIVER---TPFSSSSSS---SKHSS- 2772
 Db 2211 VLEAYRDKKLEABERTKSSISSEVDFHPPSSVIALPKWEQLSTYITSGTFQCEPQASQ 2270
 Qy 2773 -PGSTVAARVTPNPNYSPKSSADTSARPSQIPTP 2808
 Db 2271 LPSPGSPVSLPAKPRPFRRSSSLD-ISAQPPPPPP 2306

RESULT 13
 CENF_HUMAN STANDARD; PRT; 3210 AA.
 AC P49454; Q13171; Q13246;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH
 antigen).
 GN CENPF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=95379848; PubMed=7651420;
 RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
 Jones D., Yang-Peng T.L., Lee W.-H.;
 RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
 that is specifically involved in mitotic-phase progression.";
 RL Mol. Cell. Biol. 15:5017-5029(1995).
 RN [3]
 RP SEQUENCE OF 2194-3210 FROM N.A.
 RX MEDLINE=95336446; PubMed=7612011;
 RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsker T.A. Jr., Joshi H.C.;
 RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
 domain sufficient for nuclear localization.";
 RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=95370296; PubMed=7642639;
 RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
 RT "The C terminus of mitosis is essential for its nuclear localization,
 centromere/kinetochore targeting, and dimerization.";
 RL J. Biol. Chem. 270:19545-19550(1995).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=97633420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-F reveals
 interactions with the kinetochore proteins CENP-F and hBUBR1.";
 RL J. Cell Biol. 143:49-63(1998).
 RN [6]
 RP FARNESYLATION.
 RX MEDLINE=20459117; PubMed=10852915;
 RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
 Bishop W.R., Kirschmeier P.;
 RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
 and CENP-F and alter the association of CENP-E with the
 microtubules.";
 RL J. Biol. Chem. 275:30451-30457(2000).
 CC -!- FUNCTION: Probably required for kinetochore function, involved in
 chromosome segregation during mitosis. Interacts with

CC retinoblastoma protein (RB), CENP-E and BUBR1.
 CC -!- SUBUNIT: Homo- or heterodimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
 CC reorganization to the kinetochore/centromere (coronal surface of
 CC the outer plate) and the spindle during mitosis.
 CC -!- DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
 CC -!- PTM: Hyperphosphorylated during mitosis.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sb-sib.ch).
 CC
 CC EMBL; U19769; AAA82889.1; -;
 CC EMBL; U30872; AAA82935.1; -;
 CC EMBL; U25725; AAA86889.1; -;
 CC FIR; PC4035; PC4035.
 CC Genew; HGNC:1857; CENPF.
 CC GK; P49454; -;
 CC MIM; 600236; -;
 CC GO; GO:0005699; C:kinetochore; TAS.
 CC GO; GO:0005634; C:nucleus; TAS.
 CC GO; GO:0005819; C:spindle; TAS.
 CC GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
 CC GO; GO:0007088; P:regulation of mitosis; TAS.
 CC Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
 KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat; Polymorphism;
 KW Lipoprotein; Prenylation.
 FT DOMAIN 14 197 COILED COIL (POTENTIAL).
 FT DOMAIN 273 769 COILED COIL (POTENTIAL).
 FT DOMAIN 823 1328 COILED COIL (POTENTIAL).
 FT DOMAIN 1642 1746 COILED COIL (POTENTIAL).
 FT DOMAIN 1862 2987 COILED COIL (POTENTIAL).
 FT DOMAIN 2207 2568 2 X 177 AA TANDEM REPEATS.
 FT REPEAT 2207 2386
 FT REPEAT 2389 2568
 FT DOMAIN 3015 3032
 FT LIPID 3207 3207
 FT VARIANT 3202 3202
 FT CONFLICT 16 16 T -> A (IN REF. 2).
 FT CONFLICT 250 250 L -> Q (IN REF. 2).
 FT CONFLICT 272 272 G -> D (IN REF. 2).
 FT CONFLICT 611 611 MISSING (IN REF. 2).
 FT CONFLICT 1494 1589 MISSING (IN REF. 2).
 FT CONFLICT 1611 1611 V -> A (IN REF. 2).
 FT CONFLICT 1811 1811 V -> L (IN REF. 2).
 FT CONFLICT 2242 2243 ER -> DG (IN REF. 3).
 FT CONFLICT 2335 2335 L -> Q (IN REF. 3).
 FT CONFLICT 2492 2492 D -> N (IN REF. 2).
 FT CONFLICT 2545 2561 ELNVAALHNDQEAACK -> SSNRWQPCIMTKPVVS
 FT CONFLICT (IN REF. 3).
 SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
 Query Match 2.6%; Score 385.5; DB 1; Length 3210;
 Best Local Similarity 17.9%; Pred. No. 6.6e-07;
 Matches 577; Conservative 502; Mismatches 1239; Indels 897; Gaps 139;
 Qy 7 DQLKQVEALKMNSLRQLEDNSNHLTKLETASNKKEVLKQLQGSIEDEMASSGQI 66
 Db NELRRSMEENKKNLLKSHSEQKREVCHLEALKNIKQLNQSNQFAEEMKAKNTQS 534
 Qy 67 DLLERLKE-LNLDSNFPVGVKLSKSLRSYSGREGSVSSRGECSPVPMGSPRGFVN 125
 Db 535 TMLRDQEKINQENSLTLEK--KLAVADL-----EKQDCQDLI----- 574
 Qy 126 GSRESTGYLEEL-----EKESLLADLDKEKEKDWYQALQNLTKRIDSLPTENF 178
 Db 575 --KKREHHIQLNDLSKTEKESKALLSLELKKE---YELKE-EKTLFSCWSENE 627

Db	1411	FLSLOSEKILHDOHCQMSKMSLQTYVDSLKXENLVSTNLNFGODLVKEMQGLBEE	1470
Qy	1243	---POKAATC-----KVSSINQETIOTYCVEDTPICFSRSSSSLSLSASDEIGCNOTTQ	1294
Db	1471	GLVPELSSCCVPDSSSLGSDSSFYRALLEQT-----GDMSSLNLEGAVSANQCSV	1523
Qy	1295	EADSANTIQ--IAEIKGKIGIRSAE-----DPVSE-----VPAVSQHPRTKSSRL	1337
Db	1524	DEVFCSSJQTYVDSLKXENLVSTNLNFGODLVKEMQGLBEGLVPSLSSCCVPDSSSL	1583
Qy	1338	QGSLSUSSBSARHKAVE-----FPGSAKSPKSGAQTPKSP	1372
Db	1584	---SSLGSDSSFYRALLEQTGDMSSLNLEGVVSANQCSVDEVFCSSLOEBENLTRKETSPAP	1641
Qy	1373	PEHYQETPLMFSRCTSVSSL--DSFESRSIASSVQSPCSCMVSGIISPDLPSPOQTM	1431
Db	1642	AKGVSELSCEVVRQSLKUEKXESOGIMKNKEIOLEBOLLS-----SEQEL	1691
Qy	1432	PPSRKTPPPPPQTAQ--TKREVFNKAPTAEKRESGPKQAANVAQVRQVLP--DAD	1486
Db	1692	DCLRKQYLSENEQWQKLTSTVLEWESKLAEEKQT--EQLSLEVARLOQLGLDLSR	1749
Qy	1487	TLLHFATSTPDGFCSSLSLSALSDEPFIOKVELRIMPPVQVNDNGNGETESQPK--ES	1545
Db	1750	SLUGDITDADAQGRNESCDIS-----KEHSETTERTPKHVV	1786
Qy	1546	NEHQEAEKTIIDSEKDLDDSDDDIDILEECIISAMP--TKSSRGKPKQAQATASKLPPP	1604
Db	1787	HOICDKDAQQ--DNLNDIEKITECAVPTCEGSGEQSPDINYPPEGDKTQGSSECISE	1844
Qy	1605	VA-RKPSQLPVYKLLPSONRLOP--QKVSVFTPGDDMPRVYCVETGTPINFSTATSLDLTI	1662
Db	1845	LSGSPNALVPMDFLGNQEDIHNLQVRKETSNEENLRLHVIED-----RDRKV	1893
Qy	1663	EGSPNEL-----AAGEVGRGCAQSGE-----F	1684
Db	1894	ESLLNEMKELSKLHQEVOLMTKIEACIELEKIVGELKXENSLSEKLEVPFCDHQELL	1953
Qy	1685	EKRDTIPTGRSTD--BAQGGKTSSVTIPELDNKAEBGDILAE---CINSAMPKGS---	1737
Db	1954	QRVET--SEGLNSDLEMHADKSSR--EDIGNVAKVNDSKERFLDVENELSRIRSEKA	2008
Qy	1738	---HKPFRVKIMDQVQ-----QASASSAPNKNQOLDGK---KKK	1771
Db	2009	STEHEALYLEADLEVVOYETKLEKONENKOKIVCLLEEELSVTSENQRJRGELDTMSK	2068
Qy	1772	PTSPVKPIPONTETRYTRVRKNADSKNNINAEVFSNDKSKQNLKNNSKDFNDKLPNNE	1831
Db	2069	KITALDQLSEKKKEXTQLESHQSE--CLHCQVAAEVKEKTELLQTLTSSDVSELLK--	2124
Qy	1832	DRVRSFADSPHHYPTIEGTPYCFSRNDSLSLDFDDDDVDLSREKAELKAKENKESSE	1891
Db	2125	-----DKTH-----LOBKQSLSEKQALSLTKELENQIAQLNKEKE	2162
Qy	1892	AKVTSHTEL-----TSNOQSANKTOAIKQPINRGOKPILOKQSTFPQSSKPIDPRG--	1944
Db	2163	LLVKSESLOARLSRSDYEKLVNKALEALVEKGE--FALELST--QEEVHQLRRGIE	2218
Qy	1945	-----AATDEKLONFATENPTVCFSHENSSL--SSLSIDIOE--NNKVENP--IKETEPDPS	1995
Db	2219	KLVRIEADKKQALHIAEKLKERENDSLDKVENLERELQMSFENQELVILDAENSKA	2278
Qy	1996	QCEPSKQASGVAP--KSPHVEDTPVCFSRNSSLSSLSIDSDDL--LOECISS--AMPK	2049
Db	2279	EVTLTKQIEEWARSLKVPFEL--DLVTLSKENLTKQIQEKGQGLSELDKLLSFPKSLLE	2337
Qy	2050	KKKFSLKGDNEKHFPRNMGIGLBDLTLDLKDIOQRPDSEHGLSPDSENFOWKAIQEGAN	2109
Db	2338	EKEQAEIQIKESKTAVER-----LQNLKELN--EAVAALCGDQETIM--KATEQSLD	2386
Qy	2110	SIYSSLSHQAAA-----ACLSQASDSDSLSLSKSGI--SLGSPPHLATPD	2153

Db 2387 PPIEEHQLRNSIEKLARLEADEKKQLCVLQQLKESEHHDLLKGRVENLERELEIART 2446
Qy 2154 QEEKPFSTNGKPRILKGEKSTLTETKIESEKIKGKKYKYSILITGKVRNSIEGQM 2213
Db 2447 NOEHAALAEANSK-----GEVETLKA-KIEGTQSLRG---LELDVVT--IRSEKE---DL 2493
Qy 2214 KQPLQANMPSISRGRTMIHPGVNNSSTSPVSKGPPKLTSPASKSPSGQATTSPRG 2273
Db 2494 TNEIQKQERISELE-----IINSPENILQKEQKVKOMKEKST----- 2534
Qy 2274 AKPSVKSELSVAROTQIOGSSKAPRSRSDSTPSRPAQPLSRPIQSPGRNSISPR 2333
Db 2535 AMEVLQTLQKELNVAAL-----HNDQEAACKAEQNLSSQVECLELEKA--- 2579
Qy 2334 NGISPPNKLSPRTSPSTASTKSSGKMSYTSPPGMSQQNLTKQTGLSKNASSIPR 2393
Db 2580 -----QLQGLDEAKNNIVLQSSVNGLIQVEDGKQLEK---KDEISRLKNQIQD 2629
Qy 2394 SESASKGLNQ-----NNGANGANKVELGRMSSTKSSGESDRSERPVLVRQSTFI 2444
Db 2630 QEQLVSKLSQVEGBHQLWKEQNLRLNLTVELEQKIQVLQSKNASLQDTLEVL--QSSY- 2686
Qy 2445 KEASPTLRKLSESAFESLSPSRPASPTRSQATPVLSPDPMSLSTHSSVQ--AG 2502
Db 2687 -----KNLENEBELTKMDKMSFVEKYNKMTAKETELQREMHMAQKTAELOBELSG 2737
Qy 2503 GWRKLPNLSPTIEYNDGRPAKRDHDIARSHSESPSRL-----PINRSGTWKRE----- 2550
Db 2738 EKRELAGEQLLLEIKSSQDLKELTLENSLKKSLDCHMKDOVEKGVREIAYQL 2797
Qy 2551 -----HSHKSSSL-----PRVSTWR-RTGSSSSILSSSSSSSKAKSEDEKHNVSIS 2596
Db 2798 RLHEAEKKHQALLDITNKQYEVETQYREKLTSEKLECLSKLEIDLLKSSKEELNNSLK 2857
Qy 2597 GTKQ-----SKENQVSAKGTW-----RKIKENEFSPIN 2624
Db 2858 ATTQILBELKTKMDNLKYVNLKENE-RAQGNKLLIKSKQLBEEKEILOKLSQLQ 2916
Qy 2625 STSQTSSGA--TNGAESKTLIYQMAPV--SKTEDVWV-----RIED----- 2663
Db 2917 AAQEKOKTGTMTDKVDELATEIKELKETELEKTEADEYLDKYCSILLISHEKLEKAKEM 2976
Qy 2664 -----CPINPRSGR-----SPTGNTPTPIDSVSEK--ANPNTKSKDNQAKQNV-- 2706
Db 2977 LETQVHLCSQQQDGRGSLGVLGVPSPSPVTEKRLSGQNKASKGKQRSGIWE 3036
Qy 2707 -GNGSVPMRTVGLNRLTSPIQV---DAPQKGTETKPGQNNP-----VVPVSETNE 2753
Db 3037 NGGPTPATPESFSKSKKAVMSGIHPAEDTEGTEPEP-EGLPVVKKGFDIPTGKT-- 3093
Qy 2754 SP-IVETPSSSSSKHSSPSGTVAAR---VTPFNYNPSPKSKADSTS--ARPSQIP- 2806
Db 3094 SPYLLRRTTWTATSPR-----LAAQKALSPLSLGKENTLAESKPTAGGSRQSKVKV 3146
Qy 2807 ---TPVNNNTKKRDSKTDST-----ESSGTQSPK 2832
Db 3147 AQRSPVDSGTLIREPTTKSPVNNLPERSPDSDPR 3181

RESULT 14
MAPB RAT

ID MAPB RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9OW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN-Sprague-Dawley; TISSUE-Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nohias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -!- INDUCTION: By nerve growth factor.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: TO MAP1A.
CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U52950; AAB17068.1; -;
DR EMBL; X60370; CAC16162.1; -;
DR EMBL; X16623; CAA34620.1; AUT_SEQ.
DR PIR; A56577; A56577.

RESULT 15
 ID ANK3_HUMAN STANDARD; PRT; 4377 AA.
 AC Q12955;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ankyrin 3 (ANK-3) (Ankyrin G).
 GN ANK3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain stem;
 RX MEDLINE=95138209; PubMed=7836469;
 RA Kordeli E., Lambert S., Bennett V.,
 RT "Ankyring. A new ankyrin gene with neural-specific isoforms localized
 RL J. Biol. Chem. 270:2352-2359(1995).
 CC -!- FUNCTION: Membrane-cytoskeleton linker. The neural-specific
 CC isoforms may participate in the maintenance/targeting of ion
 CC channels and cell adhesion molecules at the nodes of Ranvier and
 CC axonal initial segments.
 CC -!- SUBUNIT: Neural-specific isoforms may be a constituent of a
 CC neurofascin/NRCAM/ankyrin G complex.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=A number of isoforms are produced;
 CC Name=1; Synonyms=480-kDa isoform;
 CC IsoId=Q12955-1; Sequence=displayed;
 CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues. Isoform
 CC 1 is neural-specific.
 CC -!- SIMILARITY: Contains 23 ANK repeats.
 CC -!- SIMILARITY: Contains 1 death domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U13616; AAA64834.1; --
 CC PIR; A55575; A55575.
 CC HSP; P55273; I518.
 CC Genew; HGNC:494; ANK3.
 CC
 CC MIM; 600465; --
 CC GO; GO:0006605; P:protein targeting; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000488; Death.
 CC InterPro; IPR000906; ZU5.
 CC Pfam; PF00023; ank; 24.
 CC Pfam; PF00531; death; 1.
 CC Pfam; PF00791; ZU5; 1.
 CC PRINTS; PR01415; ANKYRIN.
 CC SMART; SM00248; ANK; 21.
 CC SMART; SM00005; DEATH; 1.
 CC SMART; SM00218; ZU5; 1.
 CC PROSITE; PS50088; ANK_REPEAT; 21.
 CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PS50017; DEATH_DOMAIN; 1.
 CC Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
 KW REPEAT 73 102 ANK 1.
 FT REPEAT 106 135 ANK 2.
 FT REPEAT 139 168 ANK 3.
 FT REPEAT 172 201 ANK 4.
 FT REPEAT 203 230 ANK 5.
 FT REPEAT 234 263 ANK 6.
 FT REPEAT 267 296 ANK 7.
 FT REPEAT 300 329 ANK 8.

FT REPEAT 333 362 ANK 9.
 FT REPEAT 366 395 ANK 10.
 FT REPEAT 399 428 ANK 11.
 FT REPEAT 432 461 ANK 12.
 FT REPEAT 465 494 ANK 13.
 FT REPEAT 498 527 ANK 14.
 FT REPEAT 531 560 ANK 15.
 FT REPEAT 564 593 ANK 16.
 FT REPEAT 597 626 ANK 17.
 FT REPEAT 630 659 ANK 18.
 FT REPEAT 663 692 ANK 19.
 FT REPEAT 696 725 ANK 20.
 FT REPEAT 729 758 ANK 21.
 FT REPEAT 762 791 ANK 22.
 FT REPEAT 795 825 ANK 23.
 FT DOMAIN 1519 1898 SER-RICH.
 FT DOMAIN 4090 4174 DEATH.
 SQ SEQUENCE 4377 AA; 480399 MW; F42379E55768B684 CRC64;
 Query Match 2.6%; Score 380.5; DB 1; Length 4377;
 Best Local Similarity 18.3%; Pred. No. 1.4e-06;
 Matches 595; Conservative 430; Mismatches 1178; Indels 1051; Gaps 147;
 QY 39 TEASNMKEVLKQLOQSIEDAMASSGOIDLLERLKEINLDSSNFPVGVKLSKMS-----92
 DB 1695 TWASLSFPVKMPG--HAEVALVNGSISPLKYASSSTLINGCKATATLOEKISSATNSV 1752
 QY 93 LRSYSGREGSV-----SSRSGEC 110
 DB 1753 SSVSAATDTVEKVFSTTTAMPFPLASVSAADSAFQSLRTPSASALYTSLSGSSISATT 1812
 QY 111 SPVPMG--SPRRGFVNGSRSTCYLSELEKERSLLADLDKEKEKDYWYAOQLNITKR 168
 DB 1813 SSVTSIIITVPYVSVNVLPEPA-----LKK-----LPDSNFTKSAALLSPIKITTE 1862
 QY 169 IDSLPLETFNFSLOTLTRRQLEYEARTVRAMEPQLGTCQDMKRAQRRIARIQIIEKDI 228
 DB 1863 THQPQ---HFRTSSPVKSSLFAPSAKLSTPSSLSQSQEIILK-----DVAEMKEDL 1912
 QY 229 LRIRQLLOSQATEERSQNKHETGSHDAER-----QNEGQVGEI---NWATSGNG 277
 DB 1913 NMRTAILQTDVPEKPFQPELPKEGRIDDEEPPKIVKVKEDLVKVEIILKDKVCDVNDKG 1972
 QY 278 QGSTTRMD-----HETASVLSSTSHSAPRRLTSHLGTKEVNVYLLS 320
 DB 1973 SPKSPKDKGHPEDDWIEFSSSEIREARQQAASQSPSLPERVQVKAASEKDYNLTK 2032
 QY 321 MCGTHDKDDMSRTLLAMSSQSDSCISMRQSG---CLPLLIQLLHGNDKDSVLLGNRGS 376
 DB 2033 VIDYLTNDIGSSSLTNLYKRFDAKDGEGQKRVLPALQSHKLMPPASPVRTTSE 2092
 QY 377 KEARARASAL-HNIHSQPDKRGREIRVLHLEIQIRAYCETCWEWQEAHEPGMDQDK 435
 DB 2093 KELCNADSPFGTDTILESDDD-----FSQH-----DQDK 2122
 QY 436 NPM-----PAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGL 472
 DB 2123 SPLSDSGFETRSEKTPSAPOSAETTPKPLFHEV-PIPPVITETRTTEVVH-----2171
 QY 473 QAIARELLQVDCMYGLNDHYSTITREYAGWA-----LTNLTGADVANKAT 518
 DB 2172 -----VIRSYDPSAGDVPTQPEEPVSPKPSPTFELEPKPT 2208
 QY 519 LCSMKGMCRALVAQLKSESEDLOQVIASVLRNLSWRADVNSKTLREVSGVKALMECALE 578
 DB 2209 TSSKEKVKAFQMKASSEEDHNRVLSKGM-----2239
 QY 579 VKKESTLKSVLKSLNLSAHTENKADICAVDGAFLVGLTYRQNTNLAIIESGGGI 638
 DB 2240 VKKET-----HIT-----TTTMYVTHSPPG- 2259
 QY 639 LRNVSSLIATNEDHRQILRENNCLQTLLOHLKSHSLTVSNACQTLNLSARNP-KDQEA 697

Db 2260 -----GEGASERIEETMSVHDMKAFQ-----SGRDPSELKAG 2292
Qy 698 LWD-MGAVSMLKNLHSHXMIAMGSAALNLANRPAKYKDANIMSPGSSLSLHVXK 756
Db 2293 LFEHKSAYS-----POVHKSAETSQAHA-----KD-NQWKP-----KL 2326
Qy 757 QKALEAELDAOHLSTFDNIDNLSFKASHRSKORHKQSLYGDYVDFDNRHDDNRSDFNT 816
Db 2327 ERIIBVHEKNOAEPTEVI-----IRETKKHPEKEMVYQKLSRGRDNLKD--- 2374
Qy 817 GNMVLSPLNTVTLVPSSSSRGSLDSRSRSEKDRSLERERIGIGLVHPATENGTSSKR 876
Db 2375 -----FLPEKHDA-----FPCEEQOQEE 2395
Qy 877 GLQISTTAQIAKVMEEVSAIHT--SOEDRSGSTTELCVTDERNALRRSAAHT---H 931
Db 2396 EL-----TAESLPSVLESRVNTPVQEDSRPSSAQL--ISDSDYKTLKLSQHSIEYH 2449
Qy 932 SNTYNFTKSNRNTCSMPYAKLEYKRSNDSLSVSSNDGYK-----RGQWKPIESYSE 988
Db 2450 DDELSELRGE-SYRFAERMLLSEKLDVSHSDETSVTDHAGPPSELQSGSKRREKIAT 2508
Qy 989 DDESFKCS--YGQYFADLAHKHSANHMDNDGELDTPIYNS-----LKYSDEQ 1035
Db 2509 APKKEILSKIYXVSENGVGVKSKDEHFK-----VTVLHSGNVSSPKHAMWRFTEDR 2563
Qy 1036 LNSGRQSPSONERWAPKHIIDEIKQSEQRQSRNQSTTYPYTESTDDKHLKTPHFGQ 1095
Db 2564 LDRGREKLIYEDR-----VDRTVKEAEKLT-EVSQFFRDKTEKLAND----- 2604
Qy 1096 QBCVSP-YRSRANGSEINRVGNSHGINQVQSQCQEDDDYEDDKPTNYS----- 1144
Db 2605 -BLQPEKKARPNGKE-----YSSQOPTSSSPKVLUTELL 2640
Qy 1145 -----EERPTNYSIKYNEERXHVDPQIDYSLKYATDIPSSQK 1192
Db 2641 ASNDWEVWARQHPDQGPPEAKERAP---SLPSPPEKWLVSQOTEDS-KSTVEAKGSGIS 2696
Qy 1193 QSFPSKSSGSSKTEHMS-----SSENTSTPSSNAKQNLHPSQASRSGQPQKAAT 1248
Db 2697 QSKAPDGPQSGQLKQSKLSSIRLFEQGTAKSKDMSQEDR--KSDGQSR----- 2745
Qy 1249 CKVSS-NOBTIQTVCYVEDTPICFSCSSLSLSAEDGICN-----QTTQEAADSANTL 1302
Db 2746 IPVKLOESKLPYQV-----FAEKQOKADLDLPDESUSVQKDFWLTKDHAQSNEI 2799
Qy 1303 QIAEIXGKTGTSRSEDVSEPAVSOHPRTK--SSRLOGSSLSSESARHKAVEFPG--GAK 1359
Db 2800 VVND-----SGSDNVKK-----QRTSMSSKAMPDPSFEOQAKOLACHITSDLATR 2844
Qy 1360 SP-----SKSGAOTPKSPPE-----HYQETPLMFRCSTSVSSLDSFESRSIAS 1403
Db 2845 GPWDKVKFTWSSGATNNKSKQEKLSHVLVHDVRENHIGHPESKSVDOKNF--MSVTE 2902
Qy 1404 SVQSEPCSGMWGI-----ISPSDL-----PDSPGOTM-PPSR-----SKTPPPPP 1443
Db 2903 RERKLLTNGLSLSI KEMTVKPSKKVLYREYVVKEGDHPGLDQPSRRSESSAVSHIPV 2962
Qy 1444 QTAQTREVPKNAKAPAEKRESGPQKQAVNAVQVQVLPDADTLHPATESTPGFSKS 1503
Db 2963 RVADERMLSSNTPDGFCEQSPAPPHLSOKLSQSSMSKETVET-QHF--NSEDIEKVTY 3019
Qy 1504 SLSLSLSDPEFTQKDVLRIMPVQENDNGNETESEQPKENENOEKAEKTIDSEKDL 1563
Db 3020 SEISKVSKQSYV-----GLCPLEETET-SPTKSPDSLSFSPGKESPS-----SDV 3065
Qy 1564 LDSDDDDDIILEECIISAMPTKSKKXKKPAQATASKLPPVA-----RKPSQLPV 1614
Db 3066 FDHSPIDGLEKL-----APLAQTGEGKE-----IKTLPVYVSFVQVQKYEKIQQGV 3114
Qy 1615 YKLLPSQNLQPKHVSFTPGDDMPVYCVGEP-----INFSTATSLDLTIESPEN 1667
Db 3115 KLIISCEKTVQETRGTFYTTROQKPPSPQSGPEDDTLEQVSLDSSGKSPITETPSS 3174

Qy 1668 ELAAGEGVRGAQSGEFE-----KRDITPTECRSTDEAQQGKTSVITPELD 1714
Db 3175 E-----EVSYEFTSKTDPDSLIAYIPGKPSPIEVSSESEEEQAKSTS-----LK 3219
Qy 1715 DNKAEGDILAEINCAMPKQKSHKP--FRVKKI-----MDQVQOASASSAPNK- 1762
Db 3220 QTTVEETAVERMPND-VSKDNQRPKNRVAIEFPFPPPLDADQIESDKHHYLPKE 3278
Qy 1763 -----NQLDGKKK-----KPTSPVKP--IPONTEYRTRVRQADSNNNAER 1803
Db 3279 VDMIEVNLQDEHDKYQLAEPVIRVQPPSPVPPGADVSDSDDESIVQPVVKYTFKLKE 3338
Qy 1804 VFSNKKSKQKNUK--NNSKDF--NDKLPNNEDRVGSPAFSPPHYTFIEG----- 1851
Db 3339 VDDEQEKPKASAEKASNOKELESNGSKDNE--FGLGLDSPQNEIAQNGNNDQSITE 3394
Qy 1852 ----TPVCFSEN-----DSLSSLDFFDDDDVLSREKALR-KAKENKSEAAKVTSHTEL 1900
Db 3395 CSTATTAFRSHDDATEISLDGYDLODEDDGLTESDKLPIQAMEIKKDIWNTEGILKP 3454
Qy 1901 TSNOQSANKTQAIKQPINRGQPKPILQKQSTPPQSSKDIIDRGAATDEK--LQNFAIEN 1958
Db 3455 ADRSFQSKLEVIEE--GKVGPDDEKPPSKSSSEKTPDK--TDQKSGAQFTLEG 3507
Qy 1959 ----TPVCFSHNSS--SSLSDDIDQENNNKENEP-----KE 1989
Db 3508 RHPDRSVFPPTYSYKVDEEFATPFTVATKGL----DFDPWNNRGDDDEVFDSKREDE 3563
Qy 1990 TEP-----PDSQGEPSKPAQGYAPKSFHVEDTPVCFSRNSSL--SLSID 2033
Db 3564 TKPFLAVERDRSPATTPDT--TPARTPDEBTPS--EPNPPFHEGKMFEMTRSGAID 3618
Qy 2034 -SEDDLLQF-----CISAMPKPKKSRUKGDNE--KHSRNMGGILGEDTLIDL--KD 2082
Db 3619 MSKRDVFEERLQFPQGEHTSEKSGQDQEGDKSMVTATPQPSG--DTTVEITNLERN 3674
Qy 2083 IQRPDPSHGLS-PDSNFWDKAIQEGANSIVSSIHQAAAAACLSROASSDSDSILSKSG 2141
Db 3675 VEIPTVEPNSITSGE-----COEGTSS--SGSLEKSAATNT--SK--VDPKLTPIKMG 3725
Qy 2142 ISLGS-----PPLTPDQEPKPFPTNKG-----PRILKPGK----- 2173
Db 3726 ISASTMTWKKEGPEIT-DKIEAVMTSCGLENETITMISNTANSQMGVPEKHKDFOKD 3784
Qy 2174 ----STLETKKIES-----ESKIGKGGKKVYKSLITGK--VRS 2205
Db 3785 NFNNNNLDSSTIQDINIMSNIVLTHESAFTCTEKDNPKVSVSGKKT--GVLQGHCVRD 3842
Qy 2206 NSEISGOMKQPLQANMPSISRGRTMIHPGVNRSS-----SSTGSPVSKGPP--LKT PAS 2258
Db 3843 KQKVLGEOQXKE-----LIGIROKSKLPIKATSP-KDTFPPNHSNTKAS 3887
Qy 2259 KSPSEGQATTSPRGAKPSVKSELSPVARQTSQIGSSKAPSRSGRSDSTPSRPAQPLS 2318
Db 3888 KMKQVSO-----SEKALTTSSTSCVDKSRIPVKNTPRDNIA----- 3925
Qy 2319 RPIQSPGRNISTPGRNGISPPNKLSQLP-----RTSSPSTASTKSGSGKMS 2365
Db 3926 -----VRKACATQKQOPEKGAQKLPKLPVKVRSTCVTTTTTTTTTTTTTTT 3979
Qy 2366 YTSGRQMSQONLTKOT-----GLS-KNASSIPR-SESASGLNOMANGANKKVELSR 2418
Db 3980 CTVKVRSQLKEVCKHSIEYFKGISGETLKLVDLSSEKMOQSELS-----DEEE 4030
Qy 2419 MSSTKSGSEDRERPVLRQSTFIKEAPSPTRRLRRLKERSASPESISRRSPASP-TRS 2477
Db 4031 STSRTSLSEISRGQSVITKTSARDKKTAAPLKSEKAGEKRSRRRTGPQSPCERT 4090
Qy 2478 QAQTPVLSPLDMSLTHSSVQAGGWRKLPNPIPTIEYNDGPAPKRDHIARSHSPS 2537
Db 4091 DIRMAIVADHL--GLS-----WTELARELNFVSD-----EINQIRVENPN 4128

```

QY 2538 RLPINRS-----GTWKREHSHSSSLPRVSTWRTGSSSSILSASSESESEKAKSEDE-KHV 2592
Db 4129 SL-ISQSFMLKKWTRDGKAT-----TDALTSVLTKINRIDIVITLLEGPIDY 4177
QY 2593 NSISGKQ-SKENQVSAKGTWRKIKENEFSPNTSQTVSSGATNGAESKTLIYQMAPAV 2651
Db 4178 GNISTRSFADENNVE-----HDPVQWQNETSSG-----4207
QY 2652 SKTEDVVRIEDCPINNPRGSRPTGNTPPVIDSVSEKANPNIKDKDNQAKQNVGNSV 2711
Db 4208 -----NLESC-----AQARRVTGGLDLRLDDSPQCRDSTSYLKGEAGKFEANGS- 4253
QY 2712 PMRTVGLNRLTSFIQVDAPQKTEIKPGONNPPVSVETNES--PIVERTPFSSSSSSK 2769
Db 4254 -----HTEIT-----PEAKTSYFPBSQNDVGKQSTKETIKPKIH-----GSGH 4292
QY 2770 HSPSGTVAARVTPFNPNPRKSSADSTS-----ARPSQIPTVNNNTYKRDSDTST 2823
Db 4293 VEEFASPLAA-----YOKSLEETSKLIIETKPC---VPVSMKXMSRTSPADGK 4338
QY 2824 ESSCTQSPKRHSGS 2837
Db 4339 PRLSHEEGSSGS 4352

```

Search completed: August 25, 2004, 17:16:32
Job time : 76.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:05:26 ; Search time 183.5 Seconds
(without alignments)
4888.383 Million cell updates/sec

Title: US-09-442-489F-2
Perfect score: 14575
Sequence: 1 MAASYPQLLKQVAKWENESSGTQSPKRHSGLYLTSV 2843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvivirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10530	72.2	2829	13 P70039	P70039 xenopus lae
2	5156	35.4	1056	11 Q8BNP7	Q8BNP7 mus musculus
3	3508.5	24.1	2274	11 Q321K7	Q321K7 mus musculus
4	3414.5	23.4	2303	4 Q95996	Q95996 homo sapien
5	2229	15.3	489	11 Q8BRD8	Q8BRD8 mus musculus
6	2172.5	14.9	733	4 Q9UBZ1	Q9UBZ1 homo sapien
7	1989.5	13.7	1246	4 Q9Y632	Q9Y632 homo sapien
8	1778.5	12.2	2416	5 F91667	F91667 drosophila
9	1763	12.1	2417	5 Q9VAS9	Q9VAS9 drosophila
10	1705	11.7	1685	4 Q9UEM8	Q9UEM8 homo sapien
11	1249	8.6	324	11 Q8C493	Q8C493 mus musculus
12	1221.5	8.4	1067	5 Q9Y1T2	Q9Y1T2 drosophila
13	1220.5	8.4	1067	5 Q9G1B0	Q9G1B0 drosophila
14	1041	7.1	208	4 Q9P119	Q9P119 homo sapien
15	756	5.2	159	4 Q722Q8	Q722Q8 homo sapien
16	680	4.7	146	11 Q8C9I9	Q8C9I9 mus musculus

ALIGNMENTS

RESULT 1

P70039 PRELIMINARY; PRT; 2829 AA.
ID P70039 AC P70039; P79934;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Adenomatous polyposis coli.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
CC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Vlemnickx K., Wong B., Guger K., Gumbiner B.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64442; AAB41671.1; -;
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 5.
SQ SEQUENCE 2829 AA; 310878 MW; 8A2BABDB7706E496 CRC64;

Query Match 72.2%; Score 10530; DB 13; Length 2829;
Best Local Similarity 74.3%; Pred. No. 0;
Matches 2128; Conservative 263; Mismatches 416; Indels 58; Gaps 40;

QY 1 MAASYPQLLKQVAKWENLNRLQELDSSNHLTKLETSANWKEVYKOLQGSIEDEAM 60
DB 1 MAASYPQLKQVAKWENLNRLQELDSSNHLTKLETSANWKEVYKOLQGSIEDEAM 60
QY 61 ASSGQIDLLERLKEINLDSNFFGKVLRSKMSLRYSRGSGSVSSRGECSPVPMGSPFR 120
DB 61 ASSGPIDLLERKDLNLDSSNIPAGKARPKMSRYSRGSGSLSGHSGECSPVPGSFQR 120
QY 121 RGVNGRSTGYLELEKERSLLADLKEEKKWYQAQLNLTGYIDSLPTENFSL 180
DB 121 RGLNGRSTGYLELEKERSLLADLKEEKKWYQAQLNLTGYIDSLPTENFSL 180

QY 1021 LDTPIYSLKYSDQLNSGRQSPQSNRWARPKHIED 1058
 Db 1019 LDTPIYSLKYSDQLNSGRQSPQSNRWARPKHIED 1056

RESULT 3

QY21K7 PRELIMINARY; PRT; 2274 AA.
 AC Q921K7
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE APC2 protein.
 GN APC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99147086; PubMed=10021369;
 RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
 RA Miles A., Kuipers J., Desreux O., Peifer M., Clevers H.,
 RT "Identification of APC2, a homologue of the adenomatous polyposis coli
 RT tumour suppressor";
 RL Curr. Biol. 9:105-108(1999).
 DR EMBL; AJ130783; CAAL0207.1; JOINED.
 DR EMBL; AJ130784; CAAL0207.1; JOINED.
 DR EMBL; AJ130785; CAAL0207.1; JOINED.
 DR EMBL; AJ130786; CAAL0207.1; JOINED.
 DR EMBL; AJ130787; CAAL0207.1; JOINED.
 DR EMBL; AJ130788; CAAL0207.1; JOINED.
 DR EMBL; AJ130789; CAAL0207.1; JOINED.
 DR EMBL; AJ130790; CAAL0207.1; JOINED.
 DR EMBL; AJ130791; CAAL0207.1; JOINED.
 DR EMBL; AJ130792; CAAL0207.1; JOINED.
 DR EMBL; AJ130793; CAAL0207.1; JOINED.
 DR EMBL; AJ130794; CAAL0207.1; JOINED.
 DR EMBL; AJ130795; CAAL0207.1; JOINED.
 DR EMBL; AJ130796; CAAL0207.1; JOINED.
 DR PIR; T30258; T30258.
 DR MGI; MGI:1346052; Apc2.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR InterPro; IPR006162; Ppantne S.
 DR Pfam; PF00514; Armadillo_seg 3.
 DR SMART; SM00185; ARM; 5.
 DR PROSITE; PS00012; PHOSPHATETHEINE; 1.
 SQ SEQUENCE 2274 AA; 243137 MW; 75ABDA15D0F707F5 CRC64;

Query Match 24.1%; Score 3508.5; DB 11; Length 2274;
 Best Local Similarity 34.4%; Pred. No. 2e-164;
 Matches 1005; Conservative 356; Mismatches 789; Indels 773; Gaps 98;

QY 4 ASYDQLKVEALKMNSLRQLNSHLTKLETSNMLKQLQSGTDEA--WA 61
 Db 6 ASYEQLVQVEALKMNSLRQLNSHLTKLETSNMLKQLQSGTDEA--WA 65
 QY 62 SSGQIDLLERLKLNDSSNFFQVQLRSKMSLRYSRGSSVSSRGSSVPMGSPFRR 121
 Db 66 SSGQTEVLEQLKALQTDISSLYNLKTHAP----ALGPEP--AARTPEGSPV-HGSGPSK 117
 QY 122 -GFNGSRESTGYLEBLKERSILLADLKEBKDWYQAQLNLTPLTENPSL 180
 Db 118 DSFGELSRATIRLELDELQRCFLSEIEKEEKLWYSQGLSKRLDELPHVDTFM 177
 QY 181 QDLTRRQVEYARQIRVAMEEQLGTCDQMKRAQRRIARIQIKDIL----RIQ--- 233
 Db 178 QMDLIRQLEFEAQHRSLSWEFFGTSDEMVRQAQIRASLEQIDKELLEADRVOQTEP 237
 QY 234 --LLQCATAEKRSQNKHETGSHDAERQNEGGVGEINMA'GNGQGGTTRMDHETASV 291
 Db 238 QALLVAPVAVEEQAEPVTHPEDGTPO-----PGN----- 269

QY 292 LSSSTSHAPRLTSHLGTQKVMVYSLMLGTGDKDMSTLLAMSSQDSCISMROSQ 351
 Db 270 -----SKVEVFWLLSMLATDOEDTARTLLAMSSPSCVAMRRSG 311
 QY 352 CLPILLIQLHNDKDSV---LLGNSRSGKEPARASAAALHNIHSQPDKGRREIRVLH 408
 Db 312 CLPILLIQLHGTAGSVGRAGIPGAPGAKDARMANAALHNIIVSQPDQGLARKEMRVLH 371
 QY 409 LLEQIRAYCETCWEQEAHEPMDQKNMPAPAVEHQICPACVLMKLSFDEEHRHAWNE 468
 Db 372 VLEQIRAYCETCWDWLQARDSGTE-----TPVIEPQICQATCAVMKLSFDEEYRRANNE 426
 QY 469 LGLQIAIELLOVDCMYGLTNDHYSHITRRIYAGVALTNLTFGVDVANKATLCSMKGCWRA 528
 Db 427 LGLQIAIELLOVDCMYGLTNDHYSHITRRIYAGVALTNLTFGVDVANKATLCSMKGCWRA 486
 QY 529 LVAQLKSESEDLQOVVIAVSLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKV 588
 Db 487 IVAQLGSESEELHQVSVSLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKV 546
 QY 589 LSAWLNSAHCTENKADICAVDGALAFVLVGTLYRSQNTNLAIIESGGILRNVSLLIAT 648
 Db 547 LSAWLNSAHCTENKADICAVDGALAFVLVGTLYRSQNTNLAIIESGGILRNVSLLIAT 606
 QY 649 NEDHQILRENCLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDOEALWDMGAVMLK 708
 Db 607 REDYRQVLRDNCILQTLLOHLKSHSLTIVSNACGLTNLSARNPKDOEALWDMGAVMLK 666
 QY 709 NLIHSHKHWIANGSAAALRNLMANRPARYKDANI-MSPGSLPLSHVYKOKALEAEADAQ 767
 Db 667 NLIHSHKHWIANGSAAALRNLMANRPARYKDANI-MSPGSLPLSHVYKOKALEAEADAQ 726
 QY 768 HLSETFDNIDNLS-PKASHRSKQ-----RHQSYLYGDIYVFTNHRDNRSDNFTNMT- 820
 Db 727 HLVEHALGHEKQSLPEABTTSKPLPLRLHDG-VQDYASDSGCGFDDDDAPSLAAATTA 786
 QY 821 -----VLSPLYNTLVLPSSSSSSGSLDSRSEKDRSLERERGLGLNYHPATENPGTSS 874
 Db 787 EPASFAVMSFLPGFFLQGGALAR-----TPPARQGGLEAEKAG-----GEAAVAA 833
 QY 875 KRGLQISTTAAQIAKWEVSAIHTSQEDRSSGTTTELHCVTDERNALRRSSAAHTSNT 934
 Db 834 KAKALALAVARIDLRLVEDISALHTSSDDSGFSLSS-----GDPQAPREGRAQSCSPC 887
 QY 935 YNFT--KSENSNRTCSMPYAKLEYKRSNDLSNVSNDGVRGKGMKPSIESYSEDES 992
 Db 888 RGTGGRREAGSRAHPLRLKAAHTSLNLSNLSNGSTSDGYCTREHMT- 936
 QY 993 KFCYGYGPADLAHKIHSANHMDNDGELDTPINYSKYSDQLNSGRQSPQSNRWARP 1052
 Db 937 --CPALAL-----AEHRDD-----PVRGQ--TRP 956
 QY 1053 KHIEDIKQSEQRQSRNQSTTPVYTESTDDKHLKFPHPGQECVSPYRSRANGSET 1112
 Db 957 RRLDLDLPSRAELPARDTAATDARVT-----IKLSPTYQHVPLLD-----GAAGA-- 1002
 QY 1113 NRVSNGHINQVNSQSLCQEDYDDDKPTNYKSYRERYSEEBEERPTNYSIKYNEKRH 1172
 Db 1003 -----GVRPLVGPST----- 1013
 QY 1173 VDQPIDYSLKYATDIPSSQKQSFPSFKSSGQSKTEHMSSENSTPSSNAKRONQLH 1232
 Db 1014 -----PGARKQAM----- 1021
 QY 1233 PSSAQSRGQPKAATCKVSSINQETIQTYCVBTPICFSCSSLSLSAEDEIGCQNT 1292
 Db 1022 -IPADLSKYPEKLVASPL-PIASKVLQKVAQDGMPSLRCSLSLSLSTGHAV--PSQ 1077
 QY 1293 TQADANTSNTQIAIKG-----KIG-----TRSAEDPVSEVPVAVSQHPRTKSSLRQSSLS 1343
 Db 1078 AENLDSSSLGLEEAGPGAEGLRAWRASGTSPLFVS-IPA-----PQRGRS---GLG 1128

Db	1891	QHTQKSPVRIFFMORPARVPPPIARSPPEGSRGACAEQTPGARGRLGLVRYASAR	1950
Qy	2424	SSGESDRSERVLRVQSTFTKEAPSTLRRKLERSAGFESLPSGSRPASPTRSQOITV	2483
Db	1951	SSGES--SDRSGFRQLTTFIKESFG-LIIRRRSELGADSTASTSQAASPRGRPALPA	2007
Qy	2484	L--SPSLPDMSLSTHSSVQAGGWRKLPNLSTPTIEYNDGRPAKRHDIAHSHSEPSRLP	2540
Db	2008	VFLCSRDELVSRVPRQPLAQRSPQAKPLAPL-----APR-----RTSESERL	2055
Qy	2541	INRSGTWKREHSHSSSLFRVSTWRTGSSSSILSASSESEKAKSEDEKHVNSIS--GT	2598
Db	2056	V-RASGPRPETVKRYASLPHISVSRSDSAVSPTTQANATRRGSDGEARPLPRVAPPGT	2114
Qy	2599	KOSKENQVSAKGTWKIKENEF-----PTNSTQTVSSGATNGAESKTLIYQWAPAV	2651
Db	2115	-----TWRIKDEDVPHILRSTLPATLPLRVSSPEDSPAGTPQ-----	2153
Qy	2652	SXTEDVWRIEDCPINNPGRSPT--GNTPPVIDSVSEKANPNIKDSKNOAKQNVGNG	2709
Db	2154	RITSDAVQGTEDVAISKTNSSTSPLSESDPP-----QAP-----A	2189
Qy	2710	SVPMTKTVGLNRLTFSIQVDAPDQKGTETKPCQNNPVPVSETNESPIVERTPSSSSSK	2769
Db	2190	SGPVAPOQSGD-----VDGP-----VLTKPPASAFP--HEGLSAVIAGFP-----TSR	2230
Qy	2770	HSSPGTVAARVTPENYNSPKSS--ADSTGARPQIPTPVN	2810
Db	2231	HGSPSR--AARVPPFNYVSPMAATWASDSAVEKAPVSPAS	2271

RESULT 4

O95996 PRELIMINARY; PRT; 2303 AA.

AC O95996; TISSUE=Brain;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE DE APCL protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

RN [1]

RN SEQUENCE FROM N.A.

RP Koyama K., Nakagawa H., Nakamura Y.;

RT "APCL exon14.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RN SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=99040663; PubMed=9823329;

RA Nakagawa H., Murata Y., Koyama K., Fujiyama A., Miyoshi Y., Monden M.,

RA Akiyama T., Nakamura Y.;

RT "Identification of a brain-specific APC homologue, APCL, and its

RT interaction with beta-catenin.";

RL Cancer Res. 58:5176-5181(1998)."

DR ENBL; AB022529; BAA75469.1; JOINED.

DR ENBL; AB022518; BAA75469.1; JOINED.

DR ENBL; AB022519; BAA75469.1; JOINED.

DR ENBL; AB022520; BAA75469.1; JOINED.

DR ENBL; AB022521; BAA75469.1; JOINED.

DR ENBL; AB022522; BAA75469.1; JOINED.

DR ENBL; AB022523; BAA75469.1; JOINED.

DR ENBL; AB022524; BAA75469.1; JOINED.

DR ENBL; AB022525; BAA75469.1; JOINED.

DR ENBL; AB022526; BAA75469.1; JOINED.

DR ENBL; AB022527; BAA75469.1; JOINED.

DR ENBL; AB022528; BAA75469.1; JOINED.

DR ENBL; AB012162; BAA34611.1; -.

DR HSSP; Q02248; 3BCT.

DE Adenomatosis polyposis coli.
 GN APC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AK045053; BAC32198.1; -.
 DR MGD; MGI:88039; Apc.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0008013; F:beta-catenin binding; IDA.
 DR GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
 DR GO; GO:0009798; P:axis specification; IMP.
 DR GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
 DR GO; GO:0016055; P:wnt receptor signaling pathway; IDA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR SMART; SM00185; ARM; 1.
 SQ SEQUENCE 489 AA; 55218 MW; F7C994968B01993 CRC64;

Query Match 15.3%; Score 2229; DB 11; Length 489;
 Best Local Similarity 92.8%; Pred. No. 3.8e-102;
 Matches 440; Conservative 12; Mismatches 20; Indels 2; Gaps 2;

Qy 1 MAASVDLLKQVEALKMENSNIROLEDNSNHLTKLETSANMKVVKLOQSSIDEAM 60
 Db 1 MAASVDLLKQVEALKMENSNIROLEDNSNHLTKLETSANMKVVKLOQSSIDEAM 60
 Qy 61 ASSGQIDLLERLAKELNLDSSNFPVKLRKMSLRSYSGREGSVSSRSGECSVPVMSGSPR 120
 Db 61 -TSQIDLLERLAKELNLDSSNFPVKLRKMSLRSYSGREGSVSSRSGECSVPVMSGSPR 118
 Qy 121 RGVNRSRGSTGLVELEKERSILLADLKEEKEKWYIAQLONLTKRIDSILPTENFSL 180
 Db 119 RTFVNSRGSTGLVELEKERSILLADLKEEKEKWYIAQLONLTKRIDSILPTENFSL 178
 Qy 181 QTDLTRQLYEARQIRVAMEEQLGTCQDMKEKAQRARIQOIEKDILRILQLOSQAT 240
 Db 179 QTDLTRQLYEARQIRVAMEEQLGTCQDMKEKAQRARIQOIEKDILRILQLOSQAA 238
 Qy 241 EAERSSQKHETGSHDAERQNEGVGEINMATSGNGQGSTTRMDHETASVLSSTHSA 300
 Db 239 EAERSSQSRHDAASHEAGRQHEGHEGVAESNTAASSSQSPATRVHETASVLSSTHSA 298
 Qy 301 PRRLTSLGTVKVENVYLLSLVGLTHDKDMSRTLLAMSSQDSCISMROSCCLPLLIQLL 360
 Db 299 PRRLTSLGTVKVENVYLLSLVGLTHDKDMSRTLLAMSSQDSCISMROSCCLPLLIQLL 358
 Qy 361 HGNDKDSVLLGNRSGSKAARASAAALHNTIHSQDDPKRGRREIRVLHLEQIRAYCETC 420
 Db 359 HGNDKDSVLLGNRSGSKAARASAAALHNTIHSQDDPKRGRREIRVLHLEQIRAYCETC 418
 Qy 421 WEMQEAHEPGMDQDNPMAPVEHICPAVCVLMKLSFDEHHRHNNELGGLQA 474
 Db 419 WEMQEAHEPGMDQDNPMAPVEHICPAVCVLMKLSFDEHHRHNNELGGLKA 472

RESULT 6
 Q9UBZ1 PRELIMINARY; PRT; 733 AA.
 AC Q9UBZ1;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Qy 1931 STFPQSKDIPDCAATDEKLO-NFAIENTPVCFSHNSLSLSLSDIDQENNNKENEPK 1989
 Db 1546 KEAPAPSKAAP--AAPPAKTOPSLIADETPCVLSLSSASSLS-----EPEP--- 1591
 Qy 1990 TEPPDSQGEPSKPOASGYAPKSFHEVTPVCFNSNLSLSLSDIDDLLOECISSMPK 2049
 Db 1592 SEPPAVHPRGEPAVT-----KDPGPGGGRDSSPSP---RAABELLQRCISSALPR 1639
 Qy 2050 KKKP-SRLKCDNEKHSRNMGILGEDLTLDLKDIDQRP---DSRHG-----LSPDSEN 2098
 Db 1640 RPPVPSGLR---RRKPP-----ERPAGSRERGEAAGSDRASLDLS 1683
 Qy 2099 FDWKAIOEGANSIVSSLHQAAAAACLRQASSDSLSLSKSGISLGS-----PFHLPPDQ 2154
 Db 1684 VEMRAIOEGANSIVTWLHQAAAA--TFEASSSESLSLFSVSGLSVGTSLQPPKHKRGRQ 1740
 Qy 2155 BEKFTSNKGRILKPKGEKSTLETKKTESKKGKGGKVKYLSLITGKVRNSHISQMK 2214
 Db 1741 AEGEMGSARR-----EKGAASVKTGSPRSPAGPEK-----PRGT 1777
 Qy 2215 QPLQANMPSISGRRTMHI-----PGVRNSSSSSTSPVSKKGPPIKTPA--SKS 2260
 Db 1778 QKTFGCVAVLRGRTIVVPSFAPAPAKQKGTGPRATPRKVAP-----PCLAQAPAAKPV 1832
 Qy 2261 PSEGQTATTS-PRGAKPVKSGELSPVARQTSQIGSSKAPRSGRSDSTPSRPAQQLSR 2319
 Db 1833 PSPGQORSRLHRPAKTSSELATLSQPPRSATPPARLAKTPSSSSSOTS-----PASQPLPR 1888
 Qy 2320 ---PLOSGRNISGRNGISPPNKLSOLPTSPSTASTSGSGKMSVYSPGRQMSQ 2376
 Db 1889 KRPPVTQAA-----GALPGQASVPVTKPTALLAKHK-----TQSPVRIFFMQ 1934
 Qy 2377 NLTKGTGLSKNASSIPRESASKGLNQMNGNKA-NKKVELSRMSSTKSGSGESDRSERP 2435
 Db 1935 RPARR-GPPLARAVP--EPGPRGAGT-EAGFGARGGLGLVVRVASALSGSES--SDRS 1989
 Qy 2436 VLVROSTIKEAPSTLARKLEESASFESLSPSPSPASPTRSQAQTVLSPLDMSLST 2495
 Db 1990 GFRQQTTFIKE--SPLRRRSELSAESAASAPQAGSPRRG-----PALFAVFLCS 2040
 Qy 2496 HSSVQ-----AGGWKRLPNLSFTIYNDGRPAKHDIARSHSESFSLPILNRSGTW 2547
 Db 2041 SKCELRAPROGPAPARQPPAARPS---PGERPAR-----RTTSESPSLPV-RAPAA 2091
 Qy 2548 KREHSHKSSLLPRVSTWRTTSGSSILSASSSEKASEKBEKHVNSISGTRKQENQVS 2607
 Db 2092 RPETVKRVASLPHISVARRPDGAVPAFASADARRSSDGEPRPL-----PRVA 2140
 Qy 2608 AKG-TWRKIKENEF-----SPTNSTQTVSSGATNGAESKTLIYQMAFAVSKTEDVWVRI 2661
 Db 2141 AFGTTWRRIRDVPHILASTIPALPLRGSTPEDAPA-----GPPPKTSDAVVQT 2193
 Qy 2662 EDCPINNPRSGRSPGNT--PPVIDSVSEKANPNIKSKNOAKONVGNVGVPMRTVGLE 2719
 Db 2194 EEAAPKNTSSTPSLETRPPGAPAGGO-----LSSLGSD 2229
 Qy 2720 NLTSTFIQVADPDQKTEIKPQNNPVPVSETNESPIVER---TPFSSSSSKSHSPSGT 2776
 Db 2230 -----VDGSLA-----KAPIS-----APFVHEGLGVAVGGFPASRHGSPSR- 2266
 Qy 2777 VAARVTPFNPNPSRKSADSTSA 2800
 Db 2267 -SARVPPFNVPSPMVAATDTS 2289

RESULT 5
 Q8BRD8 PRELIMINARY; PRT; 489 AA.
 AC Q8BRD8;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

[illegible]

Db 168 ----- 167
QY 359 LLHGNDKDSVLLGNSRGSKEARASAAALHNI IHSQDDKGRREIRVLHLEQIRAYCE 418
Db 168 ----- 167
QY 419 TCWEMQEAHEPGMDOKNMPAPVEHQICPACVVLMLKLSFDEEHRHAMNELGGLOAIABL 478
Db 168 ----- EL 169
QY 479 LQVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATILCSMKGCMRALVAOLKSESE 538
Db 170 LOVDYEMHKNTRDPLNAL-RYAGMTLNLTFGDVANKATILCARGGMEALVAOLADSE 228
QY 539 DLOQVIASVLRNLWRADVNSKTLREVGSKALMECALEVKVSTLKSIVLSALWNL5AH 598
Db 229 ELHQVSSGILNLSWRADINSKNLREAGSVTALVQCVRATKESITLKSIVLSALWNL5AH 288
QY 599 CTENKADICANDGALAFVGLTLYESOTNTLAIIESGGILNRVSSLIATWEDHQILRE 658
Db 289 STENKAAICQVDGALGFVLSTLYTCQNSLAIIESGGILNRVSSLIATWEDHQILRE 348
QY 659 NNCLQTLQHLKSHSLITVSNACGTLWNL5ARNPKDQBALMDGAVSMKLNLIHSHKHMI 718
Db 349 HNCLOTLLQHLTSHSLITVSNACGTLWNL5ARSARDQELLMDLGVAGVLMRLVHSHKHMI 408
QY 719 AMGSAALRNLMANPAKYK-DANTMSGSSLP5LHVHVKQKALEBAELDAQHSETFDNID 777
Db 409 AMGSAALRNLLAHPAKHQAATAVSPGSCVPSLYVRKQALEBAELDARHQALEHLE 468
QY 778 NLSPKASHRSKO-----RHKQSLGXDYVFTNRHDDNRDNF-----NTGNMTVLS 823
Db 469 KGPFAAEATKPKPLPLHLDLQADYASDSCFDDDDAPSSLAATAATGEPASPAALS 528
QY 824 PYLNTVTLFSSSSSR-----GSLDSSSEKDRSLERERGLGNYPHATENPGTSGKRGLOI 880
Db 529 LFLGSPFLOGQAQARTPPRRGKKEAKDTSGE-----AAVAKAKAKL 572
QY 881 STTAQIAKVMEEVAIHTSQDRSGSTTEHLCHVTDERNALRBSAAHTS--NTYNTF 938
Db 573 ALAVARIDOLVEDISALHTSDSDSLSS-----GDFGQEARPREGRAQSCSPCKPREG 626
QY 939 KSENNRTCSMPYAKLEYKRSSNDLSNVSSNDGYGKGQMKPSTIESYEDDESKFCSGY 998
Db 627 REAGSRAHPLRLKAAHASLNSDLSNGSASDGYCPREHM----- 667
QY 999 QYPADLAHKIHSANHMDNDGELDTPIYLSKYKDEQLNSGRQSPQNERWAPKHIED 1058
Db 668 -LPCPLA-----ALASRREDP----- 682
QY 1059 EIKQEQROSRNQSTTYPVYVTESTDDKHLKFPQHPGQECVSPYRSRGANGSETNRVGSN 1118
Db 683 -----RCQGPSPSR----- 691
QY 1119 HGINQVNSQSCQEDDYDDKPTNYSERYSEEQHEEERPNYSIKNEKRVHDQPID 1178
Db 692 -----LDDPPGQCAEPAREATSAADAVRT---IKLSPTYQHV----- 727
QY 1179 YSLKYATDIPSSQKQSFSSKSSSGSSKTEHSSSENTSTPSSNAKQNLHPSSAQ 1238
Db 728 -----PLLEGASRAEPLADPGI-----SPCARQAWL---PADH 760
QY 1239 RSQGPQKAATCKVSSINQETIQYCVEDTPICFSRCSLSLSLSAEDICGNQTTQEADS 1298
Db 761 LSKVPEKLAAPL-SVASKALQKLAQEGFLSLRCSLSLSLSA-GRPGPSEGGLDSDS 818
QY 1299 ANTLQIAE-----IKGKIGTRSAEDPVSEVPAYVQHPRTKSSRLQSSLSSESARHKAVEF 1354
Db 819 DSSLEGLEAGPIAELEDTWRAPGATSLPVAIPAPR-----RNRGGLGVEDA----- 867
QY 1355 PSGAKSPSKSGAGTPKSPPEHYVQETPLMFSRCTSVSSLDSPESRSIASSVQSEPCSGNV 1414
Db 868 -----TPSSSENHYVQETPLVLSRCSSVSSLSGSPESRSIASSIIFSEPCSGHG 914

QY 1415 SGIISPDLPDSFGQTMPPSRKTPP--PPQPAQTKREVPKNKAFTAEKRSGPKQAAY 1472
Db 915 SGIISPELPSGQTMPPSRKTPPLAPAQ-----PPEATQFSLQWBSYVVRFLD 967
QY 1473 NAAVQVQVLP---DADTLHFEATSTPDGFCSSLSALSILDEPFIQKDVLELRIMPPVQ 1529
Db 968 IADRCERCLPSLSDAGS-VRFTEKPDENFSCASSLSALHHEHYVQQDVELRLPSAC 1026
QY 1530 ENDNGNETESEQKESNENQEAETIDSEKLLDDDDDDIIELEECIISAMPTKSSR 1589
Db 1027 PERGGAGGAGLHFAGHRRRGAGAGFSPSR-----RRGQELLELLRECLGAAPARLRK 1081
QY 1590 GKXKPAQATASKLPPVVARPKPSQLPVYKLLPSQNRLOPKHVSFTPGDDMPVYCVSGTPI 1649
Db 1082 -----VASGLVP--GRRALPVFVYMLVPAPAPQ-----EDDSCTDSASGTPV 1122
QY 1650 NFSTATSLDLTIESPENELAGEVGRGAQSGEFKEDTIPTEGRST-----DEAOGGKT 1705
Db 1123 NFSASLSDELTLQGPDPGPGPAGR-----QRPTGRPTSARQAMGHRHKAGGAGR 1174
QY 1706 SSVTIPELDNNKAEEDGILAEICINSAMPKSHKPRVKKIMDQVQOQASASSAPNKNQL 1765
Db 1175 SAEQSPGAGKNRA-----GLEPLG----- 1194
QY 1766 DGKKKPTSPVKPIQPNTVEYTRVRKNADSKNNLNAERVSFNDKSKQNLKNSKDFND 1825
Db 1195 -----RPSAPA-----DKDGSKPG----- 1209
QY 1826 KLPNNEDVRGSPAFSPHHTPIETGTPYCFSRNDSL 1862
Db 1210 -----RTRGDGALQSLCLTTPTEBAVYCYGNDVS 1239
RESULT 8
P91667
ID P91667 PRELIMINARY; PRT; 2416 AA.
AC P91667;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ADENOMATOUS polyposis COLI.
OS APC OR D-APC OR CG1451.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97144426; PubMed=8990193;
RA Hayashi S., Rubinfield B., Souza B., Polakis P., Wieschaus E.,
RA Levine A.J.,
RA "A Drosophila homolog of the tumor suppressor gene adenomatous
RT polyposis coli down-regulates beta-catenin but its zygotic expression
RT is not essential for the regulation of Armadillo.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:242-247(1997).
DR EMBL; U77947; AAB41404.1; -;
DR PIR; T13825; T13825.
DR FlyBase; FBgn0015589; Apc.
DR GO; GO:0008013; F-beta-catenin binding; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 5.
DR SMART; SM00185; ARM; 5.
DR PROSITE; PS0176; ARM_REPEAT; 2.
SQ SEQUENCE 2416 AA; 261405 MW; 4DE3A10BE10E42A9 CRC64;
Query Match 12.2%; Score 1778.5; DB 5; Length 2416;
Best Local Similarity 25.5%; Pred. No. 5.8e-79;
Matches 721; Conservative 330; Mismatches 845; Indels 935; Gaps 101;
QY 243 ERSQNKHETGSHDAERQN-EGQGVGGINVATSGNGQGTTR-----MDHETASVLSSS 295

Db 129 ELREWEHRSIDRNFERSQACQOQDLDELPRNGGGSFASAGRSPRSKEPSYITLSRFLDGD 188
QY 296 STHSAPR-----RLTSH-----LGTKEVWYVYLLMLGPHDKDDVSRITLLAM 337
Db 189 APAPAPRLPKGAAMTTTDFDERYTSSAVEATLGSKEVCVYLLMLGNSDPLEMAKFFLEL 248
QY 338 SSSQDSCISNROSCCLPILLIQLHGNKDSVLLGNRSGSKEARASAAALHNIHSQPD 397
Db 249 SGNAQSCATLRSOCVLLVQMHAPND-----GEVRKAEQALHNVVSHPDE 298
QY 398 KRGRRIRVLHLEQIRAYCETCEWQEAHEPGM--DQDNKMPAPVEHQICPAVCVLMK 455
Db 299 KAGREARVLRLLDQIVDYCFXLLQSGGEAIDSDRHL-----AASLMLK 349
QY 456 LSPDEERHAWELGGIQAELLOVDCMYG-LTNDHYISITLRRVAGMALTNLTGDDVA 514
Db 350 VSPDEERHAWELGALHAIPLNVLHDAVHGPKPDQCCNSLRRVALMALTNLTGDDEN 409
QY 515 NKATLCMKGCMRALVAQLKSESEDLQOVIASVLRNLMSRADVNSKKTLEVEGSKVLM 574
Db 410 NKALLCCQKQFMEALVAQLDSAPDILLQVTVASVLRNLMSRADSNKAVLNEIGTVTALAL 469
QY 575 CALEVKKESTLKVLSALWNLASHCTENKADI CANVGALAFVGTUTYRSQTNLAIIES 634
Db 470 AAMNRSNTLKAISALWNLASHCTENKAEFCANVGALAFVGTUTYRSQTNLAIIES 529
QY 635 GGGILRVVSLIANEDHROILRENNCLQTLLOHLKSHSLTVSNACGTLWNLARNPKD 694
Db 530 AGGILRVVSHIACEYRQILRHQCNLAILLQOLKSESLLTVVNSCGTLWNLARSAD 589
QY 695 QEALWDMGAVMLKNIHSHKXIANGSAAALRNLMANPAKYK----DANIMSPG-SSL 749
Db 590 QKFLWDMGAVMLKNIHSHKXIANGSAAALRNLMANPAKYK----DANIMSPG-SSL 749
QY 750 PSLHVRKOKALEADACHLSETFDNIDNLSPKASHRSKORHKQSLGYDVPDNRHDDN 809
Db 650 PTLKAKAKALQELGERHATCDNLD----- 677
QY 810 RSDNFTGNMTVLSPLYLNTTLPSSSSSGSLDSRSEKDRSLERBERGILGNYHPATEN 869
Db 678 -----TGG-----KLDKERASSSSRRHP----- 695
QY 870 PGTSSKGLQISTITAAQIAKMEVSAIHTSQEDRSSGTTTELHCVTDERNALRRSSAAH 929
Db 696 -----APLRT-----SAMLTKSESRDSVYSAKSDCAYD--HJIRASASD 734
QY 930 THSNTYNTKSENSNRCTCMFYAKLEYKRSNDLSNSVSSDGYGKRGOMKPSIESYSD 989
Db 735 AHR-----KVXPKITDFDLE 749
QY 990 DESKFCYGOYPADLAHKLHSAHMDNDGELDTPINYSKLYSGDEQLNSGROSPSONERW 1049
Db 750 ME-----QTEATEQPIDYVYKS----- 769
QY 1050 ARPKHIEDEIKQEQORSNQSTTPVYVTESTDDKHLKFPQHFQGCVCVPSVRSGANG 1109
Db 770 -----ENAKTSTY----- 778
QY 1110 SETNRVGSNHNINONVSQSLQBEDDVEDDKPTNYSERYSEEOHEEERTNYSIKYNEE 1169
Db 779 -----QETDL--DQPTDFSLRYAE-----NQIESDLDISGPAGQ 811
QY 1170 KRVDPIDYSLKAYTDIP--SSQKQSFSSKSSGOSKTEH-----KSSSENSTPS 1223
Db 812 KSTITP-----AETPEKSEGBILLILLDDSVKCYQTEPTPVVISAASVTLDRVAA 864
QY 1224 NAKRQNLHPS--SAQSRGQPO--KAATCKVSSINQETIYCYVEDTPICFSCSSLS 1278
Db 865 KADREAEVPEVREVTSEKAPKPKLQCGSGSVYTPKPNYCEEETPGVFSYVDSLS 924
QY 1279 SLSSADEIG-CNQTQEADSANTLOIAETKIGTRSAEDPVSEVPAVSOHPRTKSEL 1337

Db 925 SL-----DESGKANQAIUGTD-----ADIKPKLEKQEEQ----- 954
QY 1338 QGSSLSUSSARHKAVERFPSPGAKSPKSGAOPPKPPPEHYVOETPLMPSRCTSVSSLSDFE 1397
Db 955 -----SOPAEQVLTKPPTQANS-----ALETPLMFSRRSSMDSLVHDP 992
QY 1398 SRSIA-----SSVQSEPCSGMVSGIISPSDLDPGPGTQMPSPSRKTPPPPTQATKREV 1452
Db 993 DVDVANDCKSKSVSD--FSRLASGVISPEIPDPTQSMPOS----- 1033
QY 1453 PKIKAPTAEKRESGPKQAANVAQVRQVLPDADTLHLPATETPDGFCSSSSLSALS 1512
Db 1034 PERNVAGSGQNVDPVPVIFASLQPLRSVFE--DDLSSFNVEHTPAQFSTATSLNLSI- 1091
QY 1513 BPFIQDKVELRIMPVQENDNGNE-----TESEQPKESNENQKE-AEKTID 1558
Db 1092 -----VDDEKAPAVMTEDNEDELLANCINMGQKPTKPTAVKSTVVNSVDVAETIR 1144
QY 1559 S-----EKD-----LLDD--SDDDDI----- 1572
Db 1145 SVCTEDTALLSKVPSNTLSVISMSSTDPKDATAGQAQVYAHQLSDDVSSNASC 1204
QY 1573 -EILEECIISAMPTKSSRKKKPAQTAASKLPPPVARK-PSQLPVYKLLPQNRLQPKHV 1630
Db 1205 GHLLQCCIRDCM-----KKPLGEATSDPIAMLRGGNELPGY--LPS----- 1244
QY 1631 SFTPGDDMPRVYCVGTPINFSTATSLDLTIES-----PNELAAAGVGVRGGAQSGBEF 1686
Db 1245 -----ADENMK-FLVEDSPCFPSVSGISLNTVGSLSVGPVQLKETE-PSADQNP 1298
QY 1687 RDTPTGRTDEAOGGKTSSVTPIPELDNKAESGDILAEC----- 1727
Db 1299 SLANRSKRRPFWQDDSLSLSDSEDDTNLLSQAIAAGCNRPKSNLGFSSNGKRSSL 1358
QY 1728 -----INSAMPKGSHPKPRVKKINDVQOASASSASPKNQ-----LDG-KK 1769
Db 1359 SSSQPIALNAATSASSINSAMTVKSKQESYSSVSDSDNDNQSKSLFELCILKGMKT 1418
QY 1770 KKP-----TSPVKPIPQNTYRT-----RVRKNADSKNNLNAER 1803
Db 1419 KEPGARAQMQEQPIVGSVSSVQSNPSLKQFDSLVPQLPSSQVQKQRHHHHHRRER 1478
QY 1804 VFSNDKSK-KQNLKN--SKDFNDKLPNNEDRVGSAFDSPHHYTPIGTYPYCSRND 1860
Db 1479 ---BRKEKLLQECINTGISKKIN-AVPKNV--LATSAAALEPCH--PMAAT-----T 1523
QY 1861 SLSLDFDDDDVDLSREKAEIRKAKENKSEAKVTSHTELTSSQOSANKQA-IAKQPI- 1918
Db 1524 SASALSTAAPDVE--QKAH--ATSNPQOQSSHTPSSHILPNPDAIATVDTARS 1577
QY 1919 -NRQOPKPILOKQSTFPQSSKDPDRGAATDEKLQNFAN-----TPVCF----- 1963
Db 1578 PNOGNGNASQNGLET-ATGSKDLSDSEDRSDESNQSFIMETVRLDSALNETCISGASEK 1636
QY 1964 -----SHNSLSLSLSDIDQENNNKENEPKETEPPDS 1995
Db 1637 HKDPDLMLKSVRLTMBFVTSABQLRSSSNHASS-----NSHKQNSNNNTWNESTCPND 1691
QY 1996 QGEPSPKQASGAPKSPHVEDTPVCFSRNSSLSLSID-----SEDDLQECISSAMPKKK 2052
Db 1692 VSFPSVQTA-----PV-----LASLSDIDEATARSILHELIETPTNEQ 1732
QY 2053 PSRLKGNK-----HSPRNMGILGEDLTLDLQIDRPDSEHGLSPDSENFDMKAIQEGA 2108
Db 1733 PESLEGDTLVNGHADSYSGSGGLNFQJ---GGQVQAGVLEFORLLFNCTASIMT 1789
QY 2109 NSIVSSLHQAAAAACLSRQASSDSD-----SILSL-----KSGI-SIGSPF--HLTP 2152
Db 1790 NSTMIAFEARALAAENLQPAATDDDTMTFTSLNSLDLDIRPPSGWESLNSQDHSQ 1849
QY 2153 DQEEKPTSKNGPRILKPGKSTLETKIBESKGIKGGKVKYSLITGVKRSNSISQ 2212
Db 1850 SSURQAMPS--KSPRFARKMFPANLVARRAUGHLAG-----SAESVNSCNDLDN 1897

Db 530 AGGILRVSHIAVCEPYRQILRQHNCIAILLQOLKSESLTVVNSCGTILNLSARSAED 589
Qy 695 QEALWDMGAVMLKNLHSHKHIAMGSAALANLWNRPAKYK---DANINSPG-SSL 749
Db 590 QKFLWDMGAVMLKSLHSHKHIAMGSAALANLWNRPAKYK---DANINSPG-SSL 749
Qy 750 PSLHVRKOKALEARLDAHLSETFDNIDNLSPRASHRSKORHKQSLGYDVPDTRHDDN 809
Db 650 PTLERAKALQOBLGERHTAETCDNLD----- 677
Qy 810 RSDNFNTGNMTVLSPLYNTTLPSSSSRGLSDSRSEKORSERLBERGIGLGNYPHATEN 869
Db 678 -----TGG----- 695
Qy 870 PGTSSKGLQISTTAAQIAKMBEVSIAHTSQEDRSSGTTTELHCVTDERNALRRSSAAH 929
Db 696 -----APLTR-----SAMLTKSRSRDSVYSKSDCAYD---HLIRASASD 734
Qy 930 THSNYFTKSENSRNTCSMPYAKLEYKRSNDLSLNSVSDNGYGRGQMKPSESYSBD 989
Db 735 AHR-----KVKPKITFDLE 749
Qy 990 DESKFCYGOVPADLAHIIHANMDNDGELDTPINYSKYSDEQLNSGRQSPSONERW 1049
Db 750 ME-----QTEATEBPIDYSVKYS----- 769
Qy 1050 ARPKHIEDEIKQSEORQSNOSTYVYTESTDDKHLKFPQHFQGOECVSPYRSRGANG 1109
Db 770 -----ENATKSTY----- 778
Qy 1110 SETNRVGNHGINQWQSQCQEDDEDDKPTNYSERYSEBQHEEBEERTNYIKYNEE 1169
Db 779 -----QETDL---DQPTDFSLRYAE---NQTESLDIDSGPAGQ 811
Qy 1170 KRHVDPIDYSLKYATDIP---SSQKQSPFSKSSGQSSKTEH-----MSSSENTSTPS 1223
Db 812 KSTITPP-----AETVPEKSGQEILLIDDSVKYCYQEDTPYVISAASVTDLRVAA 864
Qy 1224 NAKRQNLHPS---SAQSRGQPP---KAATCKVSSINQETIOTYCYVEDTPICFSCSSLS 1278
Db 865 KADABAEVKEVRBVTSEKAPKPLKLSQSGSVTPKEPINKYCEGTGPVPSRYDLS 924
Qy 1279 SLSSAEDEIG-CNQTQOESANTLQIABTKIGITSAEDPVSFPAVSQHPRTKSSRL 1337
Db 925 SL-----DESKANQAIVGTD-----ADIKPLEKEQEOE----- 954
Qy 1338 QGSLSESARHKAVERPPGAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVSSLSDFE 1397
Db 955 -----SOPAEQVLTKPTQANS-----ALETPLMFSRSSMWSLVDHP 992
Qy 1398 SRSTA-----SSVOSEPCSMGSGIISPDLPSPGCTMPPPSKTPPPPPQTAQTKREV 1452
Db 993 DVDVANDCKSSVVD-FSRLASGVISPSBIPDSPTQSMFQS----- 1033
Qy 1453 POKAPTAERKESGPKQAANVAORVOVLPDADTLHPATESTPDGFCSSLSALS- 1511
Db 1034 PRNSVAGSQNDSPVVPVPSLQPLRSVFE---DLSSFNVEHTPAQFSTATSLNSLSIV 1092
Qy 1512 -DEFFIOXVELRIMPVQENDNGNE-----TESEQPKESNENOEKE-AEKT 1556
Db 1093 DDE-----KAPASVAEEDNEDELLANCINMGMRKFTPEAVKSTVYNSEVDVAEET 1143
Qy 1557 IDS-----EKD-----LLDD---SDDDI--- 1572
Db 1144 IRSVCTEDTALLSKVFNNTNLSVIMSSTDPKDATAGQOMTAHQLSDDVSNASDCGG 1203
Qy 1573 ---EILEECIASMPTKSRKPKPAQATASKLPPVARK-PSQLPVYKLLPSQNRLOPOK 1628
Db 1204 ASGHLLQOQIRDM-----KKPLGEATSDPIAMLRGNGNELPGY---LPS----- 1245
Qy 1629 HVSTPDDMFRVYCVGEGTPINSTATSLDLATIES-----PPNELAAGEVGRGAQSGEF 1684
Db 1246 -----ADEMKN-FLVEDSPCNFSVSGLSNLTGSLVGFVQLKETE- PSSADQNPEM 1297

Qy 1685 EKRDITPTEGSTRDEAQQGKTSSVTIPELDNDKABEGDILAEC----- 1727
Db 1298 KAKPKQOQVRRPPHQDDSLSSLSIDSEDDTNLLSQAIAAGCNRPKSNLGFSSNGKRSS 1357
Qy 1728 -----INSAMPKKGSHKPFVRVKIMDOVQOASASSGAPKNQ-----LDQ-- 1767
Db 1358 SLSSQPIAINAATASASSLNSAMTVRKSOQESYSSVSDSDNDNQSKSLFELCILKGY 1417
Qy 1768 KKKXP-----TSVPKPIPONTYRT-----RVKXNADSKNNLNA 1801
Db 1418 KTKPGARAQQOQVPIVGSSSVQSNFSLKQDFLPVQLFSSGGVQKQRQHRRHHHRRER 1477
Qy 1802 ERVFSNDKSK-KNKLKN--SKDFNDKLPNNEDVRGSPAFDPHPHTPIEGTPYCFSR 1858
Db 1478 ER---ERDEKLLQECINTGISKIN-AVPKV--LATSAAALEPCH--PMAAT----- 1523
Qy 1859 NDSLSLDDDDVDLGRKAEKAKENKESAEKVTSHTELTNSQOSANKTOIAKOPI 1918
Db 1524 -TSASALSTAAPDV-----EQKAHATSNPO---QSSSTHPSSHILENPI 1563
Qy 1919 -----NRGQPKPILOKOSTPPOSSKDIPOGATDDEKLONEAEN----- 1958
Db 1564 DAIAVTDTVRSAPAPQNGNASONGLET-ATGSKOLDSEDSDESNSQFIMETWRL 1622
Qy 1959 -----TPVCF-----SHNSSLSSLSIDDOENN 1980
Db 1623 DSALNETCISGASEKHKDPDLMLKSVRLTMEFVTSABQLRSSSHNHSS-----NSHN 1677
Qy 1981 KNEPIKETEPPPOSEPKQAPGAPKSFHVEDTPVCFSRNSSLSLSID---SEDD 2037
Db 1678 NSSNNTWNESTPCNDVSPFSVQTA-----PV-----LASLSDEDAFEAR 1718
Qy 2038 LLOECISSAMKPKKPKRLKGDNEK---HSPRNMGGILQEDITLDLKIDQRPDEHGLS 2093
Db 1719 SLHELIEITPNEQOPSELEGETDVLNGHADSVSGSGGLNFOL---GGQVQNAVRL 1775
Qy 2094 PDSNFDKWAIOEGANSIVSLHQAALCLSRQSSDS---SLSL-----KS 2140
Db 1776 FQRLFNGTASINTMTIAFEARALAEALLOPAATDDTTEMTFSLNSLDNIRPPS 1835
Qy 2141 GI-SLGSFP--HLTPDQEEKPFSTKNGPRILKCEKSTLETKIESKGIKGGKVYS 2197
Db 1836 GWESLNSCYQHSQPSLSRQAMP--KSPRPAKFPANILVARRALCHLAG----- 1884
Qy 2198 LITQKVRNSNEISQOMKOP-----LOANMPSISGRTHIHIPGVNRSSSTSPSVKKGPP 2252
Db 1885 -SASVNSCNCNLLDNKPPSLMDELDSMISVDSIQSEVADGEQDCSMATTISVSNYETA 1943
Qy 2253 L-----KTPASKSPSGOTATTSPPGAKSPVSKEL----- 2282
Db 1944 ACDDQMTVLQSCDEDEDDATNDYSSAESTPKHGSTPPNRRSLTTPDKRRLTKDRPKT 2003
Qy 2283 -----SPVARQTSQI-----GGSKAPSRGSDSTSPRPAQPLSRPI----- 2321
Db 2004 YTIATSCMEAPEANETLQIBIVEAAVPVTPSPRANGRRGSAERYKTOLIECPALIQ 2063
Qy 2322 -----KSP-----GRNSISPGR 2333
Db 2064 PQDDCPSEQLSSIRAMWQOFTFITDINIGHSQSTCSTHDPEDAGSEPCDQNSETESC 2123
Qy 2334 NG-----ISPPNKLSQL-----PRTSPSTASTKSGSGKMSYTSPCROMSQONLTQOT 2382
Db 2124 DQOEPDQLPPPSIVDLRTSVVKTLEPATAVKLVRGRKPAVSVYMQSQRN----- 2178
Qy 2383 GLSKNASSIPRESASKGLNQNMNGNANKVELSRMSSTKS-----SGSEDSRSEPV- 2436
Db 2179 -----SNNAPSKKKT-SPTIAKRSVPGSGVRLPAKKKPTP 2216
Qy 2437 -----LVQSTFIKEAPSPTLRRKLEESAPESLSPSRPASPTRSOQTPVLSPSL 2488
Db 2217 PPEPAPALERQGTIFVKD-----EPINSNVQVPEVE-TK 2249

```
QY 2489 PDMSLTHSSVQAGGWRKLPNLSPTIEYNDGRPAKRHDIAHSHSESPSLRPIINRSGTWK 2548
Db 2250 FAQTSPTHTA-----SKLPTKGTGTA-----SGGSFSG-----AGSPKRIPLAPA---R 2289
QY 2549 REHSHKSSSLPRVTRWRRTGSSSSILSASSESEKAKSEKXVNSISGTQSKENQVSA 2608
Db 2290 RMTFORANTSLRAACKGHAASRVVSGVSGTTPPSRNSNLNGSSAAAKKINHAQS 2349
QY 2609 K--GHWRIKKEHNEFSPNS---TSQTVSSGATNGAESKTLIYQAPAVSKTEDVWVRIED 2663
Db 2350 RIANWKVVDKAKTKQSSNLRKTQKSSNMLNANGTKPTLLR-----SSTFD----- 2397
QY 2664 CPINNPRSGRPTGNTPFVIDSVSEK 2689
Db 2398 -----NTPSTAGGVKSK 2409

RESULT 10
QY Q9UEM8 PRELIMINARY; PRT; 1685 AA.
AC Q9UEM8,
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE APC2 protein (Fragment).
GN APC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA Van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
RT tumour suppressor."
RL Curr. Biol. 9:105-108 (1999).
DR EMBL; AJ311187; C:extracellular matrix; IEA.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR Pfam; PF00514; Armadillo_seg. 2.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
FT NON_TER 1
SQ SEQUENCE 1685 AA; 175540 MW; B9B81923F3912F77 CRC64;

Query Match 11.7%; Score 1705; DB 4; Length 1685;
Best Local Similarity 28.4%; Pred. No. 1.5e-75;
Matches 636; Conservative 271; Mismatches 672; Indels 660; Gaps 85;

QY 654 QILRENNCLQTLLOHLKSHSLTIVNAGCTLWNLARSNPDKQALWDMGAVSMLKNLIHS 713
Db 1 QVLRDHNCLQTLLOHLKSHSLTIVNAGCTLWNLARSARQELLWDI;GAVGMLRNLVHS 60
QY 714 KHKMIAMGSAALRNIMANRPAYK-DANIMSPGSLPSLHVKKOKALEABLDQHLSET 772
Db 61 KHKMIAMGSAALRNLLAHPAKHQAAATAVSPGCVPSLYVRKQRALEABLDARHLAQ 120
QY 773 PNIDNLSPKASHRSKQ-----RHQSLGYGVVDTNRHDDNRSDN-----FNTG---- 817
Db 121 LEHLEKQGPAAEATKFLPLRLHLDGLAQDYASDSCFDDDDAPSSLAANAATGEPAS 180
QY 818 -----NMTVLSPYLNTTVLPSSSSRGSLDSRSSEKDRSLERERIGLGNYPATENPOTS 873
Db 181 PAALSIFLGPPLQGALARTPTPRG--GKEAEKDTSGE-----AAVA 222
QY 874 SREGIQTSTAQIAKWEVSAITQEDRSRSGSTTELHCVTDERNALRBSAAHTHS- 932
Db 223 AKAKIALAVARIDOLVEDISALHTSDDSFSLSS-----GDPQGEAPREGRAQSCSP 276
```

```
QY 933 -NTYNTKSENSNRTCSMPYAKLEYKRSNDSINVSNDGYKRGQMKPSIESYSEDD 991
Db 277 CRPEGRREAGRAHPLRLKAAHASLNSDSASDGYCPREHM----- 324
QY 992 SKFCSYQYPADLAHKIHSANHMDDNGELDTPINYSLKYSDEQLNSGRQSPSQNERWAR 1051
Db 325 -----LPCPLA-----ALASRREDP----- 339
QY 1052 PKHIIEDIKQSEORSRNQSTITYPVYTESTDDKHLKFQPHFGQOECVPSYRSGANGSE 1111
Db 340 -----RCGQPRPS----- 348
QY 1112 TNRVGSNHGINQVQSOLQCEDDYEDDKPTNYSERYSEBEEHEERPTNYSIKYNEER 1171
Db 349 -----LDLDFGCGQAEPPAREATSADARVT---IKLSPTVQ 382
QY 1172 HYDQPIDYSLKYATDIPSSQKQSFSSKSSQSSQSSKTEHMSSESSENTSTPSSNAKQNL 1231
Db 383 HV-----PLLEGASRAGAEPFLAGPI-----SPGARQAWL 413
QY 1232 HPSSAQSRSGQPOKAATCKVSSINQETIOTYCVEDTPICFSRCSSLSLSSAABDEIGCNQ 1291
Db 414 ---PADHLSKVPEKLAAPL-SVASKALQKLAQEGFLSLRCSLSSLSLSA-GRGPSE 468
QY 1292 TQOADSANTLQIAEIKGIGKIGTRSAB-DPVSEVPAVSQHPRTKSKRLQSSLSSESARHK 1350
Db 469 GGDLDSDSDSLEGL---EAGPSEABLDSTWRAPGATSLP----- 505
QY 1351 AVEFPSCAKSPSKS---GACTPKSPPEHYVOETPLMFBSRCTSVSSLDSPESRSIASSVQS 1407
Db 506 -VAIPARRNRGRGLGVGDAITPSSSENIVQETPLVLSRCSVSSLSGSPESPIASSIPS 564
QY 1408 EPCSGMVSGIISPDLPDPSGQTMPPSRSKTPP--PPQTAQTKREVVPKNAKTAERKES 1465
Db 565 EPCSGGSGTISPELPSGQTMPPSRSKTPPLAPAQ-----PPEATQFSLOWES 617
QY 1466 GPQAAVNAARVQVLP---DADTLHTEATSTPGFSCSSLSLSLDEPIQKDEL 1522
Db 618 YVRFIDDIADCRERCLPSELDAQS-VRFTEVKEDNFENFSCASSLALHHEHVQDDVEL 676
QY 1523 RIMPPVQNDNGNETSEQPKESNENQEAETIDSEKDLDDSDDDDIIEECIIISA 1582
Db 677 RLIPSCPEGGGAGGAGLHFAHRRREBGPATGSRPR---GAADQELRLRECLGAA 732
QY 1583 MPTKSRKKGKPAQTASKLPFPVVARPSQLPVYLLPSQNLQPKHVSTPDGDDMPRVY 1642
Db 733 VPARLRK-----VASALVF--GRRALFVPTMLVPAPAPQ-----EDDSCDT 773
QY 1643 CVEGTPINRSTATSLDLTIESPPNELAAGEVGRGAQSGEPEKRTIPTEGRSTDEAOG 1702
Db 774 SAGTVPNFSAAASLDELQGPDPGPGAGR-----QRPV---GRPTBARQA 820
QY 1703 GKTSSVTIPELDDNKABEGDILAEACINSAMPKQSHKPFVVKIMDQVQOASASSAPNK 1762
Db 821 -----MGRHK-----AGGAGRSAEQ 836
QY 1763 NQLDGGK-----KKTPSPVKPIPTQTEYTRVKNADSKNNLNAERVPFSDNKSXK 1813
Db 837 SRGAKNRAGLEFLGRPPSAPA-----DKDGSK 865
QY 1814 QNLKNSKDFNEDKLPNNEDRVGSFAFDSFPHYTPTEGTPYCFPSRNDLSLSDFFDDDDVD 1873
Db 866 PG-----RTRGDGALQSLCLTPTTEAVYCFYGNDS----- 896
QY 1874 LSREKAEELKAKENKSEAKVTSHTELTNSQOAKNTQAIKQPINRGQPKPILOKQSTF 1933
Db 897 -----DREPFAAAPTPTH-----RRTSAIPR-AFTREPR---QQRKEA 930
QY 1934 POSSKIDIPRGAATDEKIQ-NPAIENTPVCFSHNSLSLSDIDQENNNKENEPIKETEP 1992
Db 931 PAFSKAAP--AAPPPARTQPSLIADETPPCYSLSSASSLS-----EPAP---SEP 976
```


RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svarskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Yu X., Bienz M.;
RT "A new *Drosophila* APC homolog concentrated in apical adhesion zones of
RT epithelial cells,"
RL Nat. Cell Biol. 0:0-0(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
RT tumour suppressor,"
RL Curr. Biol. 9:1105-1108(1999).
DR EMBL; AF003746; AAF56249.1; --
DR EMBL; AF113913; AAD40227.1; --
DR EMBL; AF091430; AAD20985.2; --
DR FlyBase; FBgn0026598; Apc2;
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 6.
DR PROSITE; PS50176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116702 MW; 20C8F5F6121888F8 CRC64;

Query Match 8.4%; Score 1221.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 6.1e-52;
Matches 429; Conservative 177; Mismatches 380; Indels 647; Gaps 50;

QY 329 DMSRTLLAWSSQDSCINRQSGCPLLIQLLHNDKDSVLGNSRGSKARARASALH 388
Db 12 ELTRNFELESLRNPETCTALRSSDCIQLLVQILHANDE-----GLSTAKKYASQALH 62

QY 389 NTHSQPDQKRGRETRVILHLEQIRAYCETC-WEMQEAHEP-GWMDQKNPAPVEHQI 446
Db 63 NIVHNNPEEKREKREVMRLLDQLDYLVCNFTLQSGGGAIEDDRHPL----- 114

QY 447 CPAVCLMKLSFDEEHRHAMNGLGLOIAELLQVDCMYG-LTNDHYSITLRRYAGNAL 505
Db 115 -AAMKLLMKASFDEEHRQTMCELGALKAPNLVHLDAHVHGAAGREQCNALRSYGLMAL 173

QY 506 TNLITGDD--VANKATLCSMKGCORALVAQLKSESDLOQVIAVLRLNSWRADVNSKTL 563
Db 174 TNLITGDDENVHNSYLCGQRQFMEVVIQNLNAPDLQVLQVLAGVLRNLSWRADRHMTIF 233

QY 564 REVGVSKALMECALEVKESTLKSVLKLSALWNLNSAHCNTENKADICAVDALFVGLTYR 623
Db 234 NELGTVTSLARAAQMNKNLTKAILKLSALWNLNSAHCNTENKADICAVDALFVGLTYR 293

QY 624 SQNTLAIETSGGILRNVSIIATNEDHROILRENCLQTLQHLKSHLTIIVSNAGCT 683
Db 294 GPSKTLKIETAGGILRNVSIIATNEDHROILRENCLQTLQHLKSHLTIIVSNAGCT 353

QY 684 LYNLSARNPKDQALWDMGAVSMLKNLTHSKHKMTAMGSAALRNLMANRPKAKYKDANIM 743
Db 354 LWNLSARCPEDQYLLIDHNAIPLLRLLISSKNMIAEGSASALKNLNFRATQELMPN-- 411

QY 744 SPGSSLPFLHVRKQKALEAELDAQHLSETFDNIDNLSPKASHRSKQHKQSLYGVDPDT 803
Db 412 GDGGSLLP----- 418

QY 804 NRHDDNRDNFNVTGNMTVLSPYLNTVLPSSSSSRGSLDSSRSEKDRSLEREGIGLGN 863
Db 419 -----LDKEAGHG----- 426

QY 864 HPATENPGTSSKRGLOISTTAAQIAKVMBEVSAIHTSQEDRSSGSTTELHCVTDERNALR 923
Db 427 -----GLPLPR----- 432

QY 924 RSSAAHTNTYNTYKXNSNRTCSMPYAKLEVKSSNDLSNLSVSDNGYGRGQMKPISI 983
Db 433 -----FSSLSSNPTGSLKKVRPS-----TVSTTGLFNKRKRESRESI 470

QY 984 ESYSEDESKEFCSGYGOYPADLAHKAHSAHMDNDGELDTPINYSKYSDEQLNSGRQSP 1043
Db 471 YSKGSD-----STKYST----- 482

QY 1044 SONERWARPKHIEBIKQSEQRNQSTTVYVYTESTDDKHLKQPHGQOECVSPYR 1103
Db 483 -----K 483

QY 1104 SRGANGSETRVGSNHGINQVSQSLCQEDDYEKKPTNYSERYSEEQHEEERPTNYS 1163
Db 484 SEG-----KNPFEIVTPT-----EQPIDYS 505

QY 1164 IKYNEEKRH-----VDQPIDYSLKYATDIPSSQKQSFSPKSSSGSSGSSKTEHSSSS 1215
Db 506 MKYWEHKPNSSKTFEIDLQDPTDPSARY-----KERRSAQTPAKELKSETIN 551

QY 1216 ENTSTSSNAKRONQLHPSSAQSR-----SQPKAATCKVSSINQEIQYCVYE 1265
Db 552 EIRS-----KELQLTKSSSATELRNSPGLVAWSAAKQKIAT-ETETAEAPINCEE 603

QY 1266 DTPICTSRCSLSLSLSAEDIGCQNTTQADANTLQIAEIKGKIGTRSAEDPVSVEPA 1325
Db 604 GTFGSFRPDSINSL-----TEKPE 623

QY 1326 VSGHPTKSSRLSGSSLSSESARHKAVEPPSGAKSPSKGAQTPKSPPEHYVQETPLMFS 1385
Db 624 KCMPPKPTPTKT-----AVLPVHVDGNTPNQIDS--ALETPLMFS 660

QY 1386 RCTSVSLDSFESRSIA-----SSVQSEPCSGMVSGIISPSDLPDPSGQTMPP--PSRSKTP 1439
Db 661 RRSMDSLVG-DDETVACEDNGSVISE-YSRMQSGVISPSELPSDPTQSPQSPRRDR-- 716

QY 1440 PPPPQTQATKREVPKNAKPTAKRESGPKQAAVNAAVQVRVQLPDADTLHLHATESTPDG 1499
Db 717 --KVSTQNNLDTPEQKPSVFE-----DKLNRHFVHEHTPAA 750

QY 1500 FSCSSLSLSALSDPEFTQKDVLRIMPPVQNDNGNETESEQPKESNENOE-KEAEKTIID 1558
Db 751 FSCATLSLSLSM-----MDDSNANAIQRGRNDINGNDGAPRSYCTED 793

QY 1559 SEKDLDDSDDDIIELECIIISAMPTKSSRGKPKAQATSKLPPVPA-----RKPSQ 1611
Db 794 TTAIVLSKAPNSDLSIL-----SIPNDIN-----ANEAPVPAEPADVTGMDTRMPAE 841

QY 1612 LPVYKLLPSQNRLOPQKHVSFTPODDMPRVYCVGEPINPSTATSLDITIEFPFNLAA 1671
Db 842 DAISKMECGGNALP-----SYLPVSDMSKYVEDSCTFSVIGLSHLTV----- 887

QY 1672 GEGVRGGAQSGEPEKRDITIEGRSTDEAOGKTSVTIPELDNKK-AEEDGDIILAEINS 1730
Db 888 -----GSAKAGPVUK---LPM--RTAEAAQ-----PKLPPRSVAVQGD-----AEP 924

QY 1731 AMPKXGSHKPRVKKIMDQ-----VQOASASSAPNKNQLDGKKKKTPSPVKPIQNTYR 1786

```

Db 925 RLPPKSDSSUSMDSDDDCNLLSQAIAAGSC-----RPOFSGA----- 964
Qy 1787 TRVEKNADSKNNNAERVFSDNCKSKQNLKNNKSFDFKLPNNEDVRGSPAFDSPHHY 1846
Db 965 -----STSSSLANASTSTLCRENGOSKKQ-----VEHGDK-PN----- 996
Qy 1847 TPIEGTYCFSRNDSLSLDFDDDDVLSREK-----AELRKAKENKESA----- 1892
Db 997 -----YSSDDSL-----DDDDDDARSKSLFEQCLLSGMHKSNDALASEGEPPGOR 1041
Qy 1893 -KVTSHTLSNQ 1904
Db 1042 QETISARDRFVSNQ 1054

RESULT 13
Q961B0 PRELIMINARY; PRT; 1067 AA.
AC Q961B0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE L024920P.
GN APC2 OR C66193.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AV051719; AAX93143.1; -.
DR FlyBase; FBgn0026598; Apc2.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 7.
DR DR PROSITE; PS0176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116678 MW; 03C4119AEF19D198 CRC64;

Query Match 8.4%; Score 1220.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 6.9e-52;
Matches 429; Conservative 177; Mismatches 380; Indels 647; Gaps 50;

Qy 329 DMSRTLLAMSSQDSICSMRQSGCLPLLQLLGHNDKDSVLLCNSGSKSEARASAAHL 388
Db 12 ELTRNFLELSRNETCTALRSSDCIQLLVQLHANDE-----GLSTAKYASQALH 62
Qy 389 NIHTSPDDKGRRETRVLHLEQIRAYCETC-WEWQEAHEP-GMDQKNPMPAVEHQI 446
Db 63 NIVHNPPEEKREVKRMFLRLDQILDYCNFLHTQLQSGGEAIDEDRHPL----- 114
Qy 447 CPACVLMKLSFOEERHMANELGGLOAJAEILOVDCEMYG-LITNDHYSITLRRYAGMAL 505
Db 115 -AAXKLLMKASFEEHROTWCCELGALKALPNLVHLDHVGGAAGWECNALRSTGLMAL 173
Qy 506 TNLTFGD--VANKATLCMSKGNRUALVAQKSESEDLOQVIAVLRLNLSWRADVNSKTL 563
Db 174 TNLTFGDENVHNSYLCGQRFQEVVIAQLNTPAPDELLQVLAVLRLNLSWRADKEMKTF 233
Qy 564 REVGSVKALMECALEVKKSTLKSVALNLSAHCTENKADICAVDGLALAVGLTYR 623
Db 234 NELGTVTSLARAQNKENTLKALUSALWNLSAHCTNKAFCVADGALAVGLMLSTE 293
Qy 624 SQNTNLAIIESGGGILRNVSLSIATNEDHRQILRENNCLQTLLOHLKSHSLTIVSNACGT 683

```

```

Db 294 GPSKTLKLIENAGGILRNVSISHAVCEPYQILRRYNCIALLOOLKSESITVVSNSCGT 353
Qy 684 LWNLSARNPKDOEALWDMGAVSMKLNLIHSHKMIAMGSAALNLANRPAKYKDANIM 743
Db 354 LWNLSARCPEDQOYLIDHNAIPLRALISSKXSMIAEGSAGALKLVNFRATQELMPE-- 411
Qy 744 SPGSLPSLHVRRKQKALEAEADLAQHLSETFDNIDNLSFKASHRSQRHKOSLYGDIYFDT 803
Db 412 GDGGSLLP----- 418
Qy 804 NRHDDNRSDNFTGNMTVLSPLYLNTTVLPSSSSRGSLDSRSSEKRSKSLERERIGLGNV 863
Db 419 -----LQKBAGH----- 426
Qy 864 HPATENPGTSSKRGLOISTTAAQIAKYMEEYSAIHTSQEDRSSTTTELHCVTDERNALR 923
Db 427 -----GTLPRR----- 432
Qy 924 RSSAAHTSHNTYNTFKSENRTCSMPYAKLEYKRSNDSNLSVSSNDGKGRQMKPDSI 983
Db 433 -----FSSLRLSSNPTGSLKKVRPS-----TVSTTGFLNRKCESEESI 470
Qy 984 BSYSEDDDESKFCSYGQYPAIDLAHKIHSANHMDDNDGELDTPINYSKYSDQLNSGRQSP 1043
Db 471 YSGKSD-----STKYST----- 482
Qy 1044 SQNERWAPKHIIEDEIKQSEQRQSRNQSTTYPVYTESTDDKHLKFPQHFQOECVSPYR 1103
Db 483 -----K 483
Qy 1104 SRGANGSETNRVSGHGINQVSQLCEDDYEDDKPTNYSERVSEBEQHEEERPTNYS 1163
Db 484 SEGA-----KNPFELVTP-----BEQPIDYS 505
Qy 1164 IKYNEEKRH-----VDQPIDYSLKYATDIPSSQKQSFSSKSSGSGQSKTEHSSSS 1215
Db 506 MKYMEHPNPSKTFEIDLDQTFDSARY-----KERRSAQTAQPELSETN 551
Qy 1216 ENTSTPSNAKRONLHPSSAQSR-----SGOPOKAATKYSSINQETIYQYCE 1265
Db 552 BIRS-----KELQLTSSSATELRNPGLVASAAKQKIAT-ETETETAERDIPNYCEE 603
Qy 1266 DTPICFGRCSLSLSLSSAEDEIGCNQTTQBADGSANTILQIAEIKGKIGTRSAEDFVSEVPA 1325
Db 604 GTPGFSRFDLSNL-----TEKPE 623
Qy 1326 VSQHPRTKSSRLQSSLSSESARHKAVERFPSSGAKSPSKGAQTPKSPPEHYVQETPLMFS 1385
Db 624 KCMPPKTPKT-----AVLPVHDGNTFQONIDS--ALETFLMFS 660
Qy 1386 RCTSVSLDSFEGERSIA-----SSVQSEPCSGMVGSIIPSDLPDPSQOTMP--PSRSKTP 1439
Db 661 RRSMDSLVG-DEETVACEDNGSVISE-YSRQSGVISPSELDPDPSQSNPQSPRRDR-- 716
Qy 1440 PPPQTAQTKREVPKNKAPTAKRESGPKQAANAARVQVQLPDDATLLHFATETPDG 1499
Db 717 ---KVSTQNLLDTPQKPFSTVFE-----DKLNRFFVEHTPAA 750
Qy 1500 FSCSSLSLSLSDPEPFTQKQDVELRIMPVQDNGNETSEQPKENQOE-KEAEKTIID 1558
Db 751 FSCATSLNSLSM-----MDDSNANAIQRQDNGDINGDAPRSCTED 793
Qy 1559 SEKDLLDDSDDDIIELEECIIISAMPTKSRKKGKPAQTAASKLPPVPA-----RKPSQ 1611
Db 794 TTAVLKRAPNSDLSIL-----SIPNDLN-----ANEAPVPAADVDTGMDTRMPAE 841
Qy 1612 LPVYKLLPSNRLQPKQKHSFTPGDMPRVYCVGEGTPINFSTATISDLTIESPPNELAA 1671
Db 842 DATSKMCGGNALP-----SYLPVDEMSEKYYVEDSCTFSVLSGLSLTV----- 887
Qy 1672 GEGVRGGAQSGEPEKRDITPTEGRSTDEAQQGKTSSTVPIPELDDNK-ABEGDILAEFINS 1730

```

Db 888 -----GSAKAGVFLK---LPM---RTAEAAQA-----PKLPRRSVAVQCD-----AEP 924
QY 1731 AMPGKSHKPRVVKINDQ-----VQASASSAPNKNQDGGKKPSPVKPIPQNTYR 1786
Db 925 RLPPKSDLSLSLSDSDCNLLSQATAAGC-----RQPSGA--- 964
QY 1787 TRVRNADSKNNLNAERVFSDKSKQNKNSKDFNDKLPNNEDRVRGSFAPDSPHY 1846
Db 965 ----STSSSLANASTSTLCRENGQSKQ-----VEHGDK-PN----- 996
QY 1847 TPIEGTPCFNRNDSLSLDFDDDDVLSREK-----AELKAKENKESAE----- 1892
Db 997 -----YSDDSL-----DDDDDDARSKSIFQCCILSGMHKSNDLASEGEPPGQR 1041
QY 1893 -KVTSHTELTGNQ 1904
Db 1042 QEISARDRVSNQ 1054
RESULT 14
ID Q9P119 PRELIMINARY; PRT; 208 AA.
AC Q9P119,
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Adenomatosis polyposis coli tumor suppressor (fragment).
GN APC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20435068; PubMed=10982189;
RA Su L.K., Steinbach G., Sawyer J.C., Hindi M., Ward P.A., Lynch P.M.;
RA "Genomic rearrangements of the APC tumor-suppressor gene in familial
RT adenomatous polyposis";
RL Hum. Genet. 106:101-107(2000).
DR EMBL; AF127506; AAF34355.1; -;
DR EMBL; AF127034; AAF34355.1; JOINED.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR00225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 2.
FT NON_TER 1
FT NON_TER 208
SQ SEQUENCE 208 AA; 22726 MW; 51076F257B61C528 CRC64;
Query Match 7.1%; Score 1041; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 5.4e-44;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 471 GLQAIALLQVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALV 530
Db 1 GLQAIALLQVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALV 60
QY 531 AQLKSESDLOQVIAVLNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSVL 590
Db 61 AQLKSESDLOQVIAVLNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSVL 120
QY 591 ALWNLSAHCNTKADICAVDGLAFVLGTLTYRSQNTLAIIESGGILRNVSLLIATNE 650
Db 121 ALWNLSAHCNTKADICAVDGLAFVLGTLTYRSQNTLAIIESGGILRNVSLLIATNE 180
QY 651 DHRQILRENNCLOTLLQHLKSHSLTI 678
Db 181 DHRQILRENNCLOTLLQHLKSHSLTI 208
RESULT 15
ID Q72208 PRELIMINARY; PRT; 159 AA.
AC Q72208;

DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Pange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards D.K., Wozney K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC056268; AAF56268.1; -;
KW Hypothetical protein.
FT NON_TER 159
SQ SEQUENCE 159 AA; 17843 MW; E0C0CC055A22C91B CRC64;
Query Match 5.2%; Score 756; DB 4; Length 159;
Best Local Similarity 98.7%; Pred. No. 4.5e-30;
Matches 153; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAASYDOLLKQVEALKVENSRLRQLEDNSNHLTKLETEASNKVEYLKQLQGSIEDEAM 60
Db 1 MAAASYDOLLKQVEALKVENSRLRQLEDNSNHLTKLETEASNKVEYLKQLQGSIEDEAM 60
QY 61 ASSQCIDLLRLKELNLDSSNPPGVKLRSMKSLRSYSGSRSRSGCSVPVPMGSFPR 120
Db 61 ASSQCIDLLRLKELNLDSSNPPGVKLRSMKSLRSYSGSRSRSGCSVPVPMGSFPR 120
QY 121 RGFVNGSRSTGYLLELEKERSLLLDLDKKEKK 155
Db 121 RGFVNGSRSTGYLLELEKERSLLLDLDKKEKK 155
Search completed: August 25, 2004, 17:22:36
Job time : 218.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:01 ; Search time 180.5 Seconds

(without alignments)
4450.319 Million cell updates/sec

Title: US-09-442-489F-2

Perfect score: 14575

Sequence: 1 MAASVQLKQVZALKWEN.....ESSGTQSPKHSGLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14575	100.0	2843	2 AAR26052	Aar26052 APC gene
2	14575	100.0	2843	2 AAW35392	Aaw35392 Human ade
3	14575	100.0	2843	2 AAW38370	Aaw38370 Human ade
4	14575	100.0	2843	5 ABG90964	Abg90964 Human ade
5	14551	99.8	2860	2 AAR63507	Aar63507 Adenomat
6	14533	99.7	2843	2 AAW76140	Aaw76140 Human APC
7	14533	99.7	2843	2 AAW76144	Aaw76144 Human APC
8	14533	99.7	2843	3 AAB23011	Aab23011 Human APC
9	14533	99.7	2843	5 ABG71105	Abg71105 Human ade
10	14533	99.7	2973	2 AAW76821	Aaw76821 Human APC
11	14533	99.7	2973	4 AAY72782	Aay72782 Transcrip
12	14526	99.7	2973	3 AAY70304	Aay70304 Protein u
13	14521	99.6	2843	2 AAR58634	Aar58634 Adenomat
14	14515.5	99.6	2842	2 AAR63508	Aar63508 Adenomat
15	14515.5	99.6	2842	5 ABG90968	Abg90968 Human APC
16	14515.5	99.6	2843	2 AAW11922	Aaw11922 Adenomat
17	14506	99.5	2843	7 ADE65846	Ade65846 Human ade
18	13966.5	95.8	2742	7 AAB23012	Aab23012 Human APC
19	13168.5	90.3	2842	7 ADS56175	Adse56175 Rat Prote
20	4719	32.4	912	5 ABG71106	Abg71106 Human ade
21	3889	26.7	767	5 ABG71107	Abg71107 Human ade
22	3508.5	24.1	2274	4 AAB50674	Aab50674 Mouse APC
23	3476.5	23.9	902	4 ABG09335	Abg09335 Novel hum
24	3414.5	23.4	2303	6 ABR58648	Abr58648 Human can
25	2171.5	14.9	799	3 AAY92061	Aay92061 Human APC

26 1767.5 12.1 1674 3 AAY92060 Murine AP
27 1763 12.1 2417 4 AAB58126 Drosophil
28 1478 10.1 332 2 AAR88353 Mutant A-
29 1221.5 8.4 1067 4 ABB62156 Drosophil
30 556.5 3.8 2344 4 AAU37120 Staphyloc
31 539 3.7 2768 4 ABB68397 Drosophil
32 526.5 3.6 2586 4 ABB66878 Drosophil
33 524.5 3.6 3257 4 ABB67502 Drosophil
34 518 3.6 2271 6 ABUL6000 Protein e
35 517.5 3.6 2271 6 ABP56876 Staphyloc
36 517.5 3.6 2283 6 ABP56876 Staphyloc
37 516.5 3.5 1186 4 AAB50654 C. elegan
38 514.5 3.5 2261 6 ABJ18914 Pathogen
39 507.5 3.5 178 2 AAW33894 Flea sal
40 507.5 3.5 178 2 AAW82368 Flea sal
41 493.5 3.4 2137 5 ABP39618 Staphyloc
42 490 3.4 3111 4 ABB60327 Drosophil
43 481.5 3.3 2951 4 ABB60291 Drosophil
44 474.5 3.3 5533 4 ABB65772 Drosophil
45 474.5 3.3 5560 4 ABB71160 Drosophil

ALIGNMENTS

RESULT 1
AAR26052
ID AAR26052 standard; protein; 2843 AA.
XX AC AAR26052;
XX DT 25-MAR-2003 (revised)
DT 28-JAN-1993 (first entry)
XX DE APC gene product in familial adenomatous polyposis.
XX neoplasm; cancer; oncogene; tumour; growth; detection; diagnosis;
XX prognosis; treatment; sporadic colorectal carcinomas; ss.
XX Homo sapiens.
XX WO9213103-A1.
XX PD 06-AUG-1992.
XX PF 16-JAN-1992; 92WO-US000376.
XX PR 16-JAN-1991; 91GB-00000963.
XX PR 08-AUG-1991; 91US-00741940.
XX FA (UYJO) UNIV JOHNS HOPKINS.
XX FA (ICIL) IMPERIAL CHEM IND PLC.
XX FA (UTAH) UNIV UTAH.
XX FA (CANC-) CANCER INST.
XX PI Kinzler KW, Vogelstein B, Anand R, Hedge PJ, Markham AF;
PI Albertsen H, Carlson ML, Groden JL, Joslyn G, Thliveris A, White RL;
PI Nakamura Y;
XX WPI; 1992-284685/34.
XX N-PSDB; AAQ27234.
PT Detection of somatic and germ-line alterations of human APC gene - used
PT to diagnose, treat and study familial adenomatous polyposis and sporadic
PT colorectal cancer.
XX Disclosure; Page 47; 132pp; English.
XX This sequence is encoded by the APC (Adenomatous Polyposis Coli) gene
XX associated with tumorigenesis, found on chromosome 5q. The sequence may
XX be mutated by deletions, insertions, inversions, or point mutations of the
XX gene. The APC gene is expressed in most normal tissues as well suggesting
XX that APC is a tumour suppressor. (Updated on 25-MAR-2003 to correct EN

CC	field.)	(Updated on 25-MAR-2003 to correct PI field.)
XX	Sequence 2843 AA;	
SQ	Query Match	100.0%; Score 14575; DB 2; Length 2843;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 2843; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	MAAASYDQLLKQVBAKXVNSNLROELEDNSNHLTKLTEASNMKVYLKQJQSIEDEAM 60
Db	1	MAAASYDQLLKQVBAKXVNSNLROELEDNSNHLTKLTEASNMKVYLKQJQSIEDEAM 60
Qy	61	ASSGQIDLLERLKEINLDSNFPVGLKRSKMSLRSYSGREGSVSRSGECSPPVMGSPFR 120
Db	61	ASSGQIDLLERLKEINLDSNFPVGLKRSKMSLRSYSGREGSVSRSGECSPPVMGSPFR 120
Qy	121	RGFVNGSRSTGYLEELKERSLLADLDKEEKEKDWYAAQLQNLTKRIDSLPTENFSL 180
Db	121	RGFVNGSRSTGYLEELKERSLLADLDKEEKEKDWYAAQLQNLTKRIDSLPTENFSL 180
Qy	181	QTDLTRQLLEYEARQIRVAMEEQIGTQDWEKRAQRRIARIQITEKOILRIQLLQSQAT 240
Db	181	QTDLTRQLLEYEARQIRVAMEEQIGTQDWEKRAQRRIARIQITEKOILRIQLLQSQAT 240
Qy	241	EAEASSQNKHETGSHDAERQNEGQGVGEINWATSGNGQGSRTTRMDHETASVLSSSSTHSA 300
Db	241	EAEASSQNKHETGSHDAERQNEGQGVGEINWATSGNGQGSRTTRMDHETASVLSSSSTHSA 300
Qy	301	PRRLTSHLGTKVEMVYSLLSMLGTHDKDDMSRTLLAVSSQDSCISMRQSGCPLLIQLL 360
Db	301	PRRLTSHLGTKVEMVYSLLSMLGTHDKDDMSRTLLAVSSQDSCISMRQSGCPLLIQLL 360
Qy	361	HGNDKDSVLLGNSRGSKEARASAAALHNIHSOPDDKGRBRIRVLHLEEQIRAYCETC 420
Db	361	HGNDKDSVLLGNSRGSKEARASAAALHNIHSOPDDKGRBRIRVLHLEEQIRAYCETC 420
Qy	421	WEMQEAHEPGVDQKNPMPAPVEHQICPAVCVLMKLSFDEHRHANNELGGLQAIABLLQ 480
Db	421	WEMQEAHEPGVDQKNPMPAPVEHQICPAVCVLMKLSFDEHRHANNELGGLQAIABLLQ 480
Qy	481	VDCEMYGLTNDHYSITILRRYAGNALTNLTGADVANKATILCSMGCKRPAVLAQLKSSEDL 540
Db	481	VDCEMYGLTNDHYSITILRRYAGNALTNLTGADVANKATILCSMGCKRPAVLAQLKSSEDL 540
Qy	541	QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKEKSTLKSVALNLSAHC 600
Db	541	QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKEKSTLKSVALNLSAHC 600
Qy	601	ENKADICAVDGAFLVGTITYRSQNTNLAIIESGGGILRNVSGLIATNEDHQIILRENN 660
Db	601	ENKADICAVDGAFLVGTITYRSQNTNLAIIESGGGILRNVSGLIATNEDHQIILRENN 660
Qy	661	CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDOEALWDGAVSMLKNLIIHSHKHWIAM 720
Db	661	CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDOEALWDGAVSMLKNLIIHSHKHWIAM 720
Qy	721	GSAAALRNLMANRPKYKDANIMSPGSSLSPLHVRKQKALEABDQHLSETFDNIDNLS 780
Db	721	GSAAALRNLMANRPKYKDANIMSPGSSLSPLHVRKQKALEABDQHLSETFDNIDNLS 780
Qy	781	PKASHRSKQRHKOSLYGDVYDFDNRHDDNDSNFTGNMTVLSPYLNNTTVLPSSSSSRGS 840
Db	781	PKASHRSKQRHKOSLYGDVYDFDNRHDDNDSNFTGNMTVLSPYLNNTTVLPSSSSSRGS 840
Qy	841	LDSSRSEKRSLERERIGLGNTHYPATENPGTSSKRGLOISTTAAQIAKWEVSVSAIHTS 900
Db	841	LDSSRSEKRSLERERIGLGNTHYPATENPGTSSKRGLOISTTAAQIAKWEVSVSAIHTS 900
Qy	901	QEDRSSGSTTELCVTDERNALRRSSAAHSTHNTYNTFKSENRRTCSPYAKLEYKRSS 960
Db	901	QEDRSSGSTTELCVTDERNALRRSSAAHSTHNTYNTFKSENRRTCSPYAKLEYKRSS 960
Qy	961	NDSLNYSVNDGKRGQMKXPSIESTSEDDSEKFCSGYQVADLAHKHISANFMDNDGE 1020

Db 2041 ECISAMPKXKPKRLKGDNEKHSFPRNMGGLGDLTLDLKDIOQPDSEHGLSPDSENF 2100
 QY 2101 WKATQEGANSIVSSIHQAAACLSRQASSDSDLSIKSGISLGSPPHLPDQEEKPF 2160
 Db 2101 WKATQEGANSIVSSIHQAAACLSRQASSDSDLSIKSGISLGSPPHLPDQEEKPF 2160
 QY 2161 SNKGRILKPEKSTLETTKIESKSGIKGKKVYKSLITGKVRNSSEISQMKQPLQAN 2220
 Db 2161 SNKGRILKPEKSTLETTKIESKSGIKGKKVYKSLITGKVRNSSEISQMKQPLQAN 2220
 QY 2221 MPSISRGTMHIHPCVRNNSSTSPVSKKGPPLKTPASKSPSEGTATTSPRGAKPVKS 2280
 Db 2221 MPSISRGTMHIHPCVRNNSSTSPVSKKGPPLKTPASKSPSEGTATTSPRGAKPVKS 2280
 QY 2281 ELSPVARTSIOIGSSKAPSGSGRSDSTPSPAQOPLSRPIQSPCRNSISPGRNGISPPN 2340
 Db 2281 ELSPVARTSIOIGSSKAPSGSGRSDSTPSPAQOPLSRPIQSPCRNSISPGRNGISPPN 2340
 QY 2341 KLSQLPRTSSPTASTKSGSGKWSYTPGQMSQOONLTQGLSKNASSIPRESASKG 2400
 Db 2341 KLSQLPRTSSPTASTKSGSGKWSYTPGQMSQOONLTQGLSKNASSIPRESASKG 2400
 QY 2401 LNMOMNGANKVLSRWSSTKSGSSESDSERPVLVRQSTFIKEAPSPTRLEKLEESA 2460
 Db 2401 LNMOMNGANKVLSRWSSTKSGSSESDSERPVLVRQSTFIKEAPSPTRLEKLEESA 2460
 QY 2461 SPESLSPSSRPASPTRSQOATPVLSPSLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Db 2461 SPESLSPSSRPASPTRSQOATPVLSPSLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 QY 2521 RPAKHDIARSHSSEPSRLPNRSGTWREHSHSSSLPRVSTWERTGSSSSILSAGES 2580
 Db 2521 RPAKHDIARSHSSEPSRLPNRSGTWREHSHSSSLPRVSTWERTGSSSSILSAGES 2580
 QY 2581 SEKAXSEDEKHVNSISGTSKQKENQVSAKGTWRKIKENEFPTNSTGTSVSGATNGAES 2640
 Db 2581 SEKAXSEDEKHVNSISGTSKQKENQVSAKGTWRKIKENEFPTNSTGTSVSGATNGAES 2640
 QY 2641 KTLIYOMAPAVSKTDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKAMPNIKSDKN 2700
 Db 2641 KTLIYOMAPAVSKTDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKAMPNIKSDKN 2700
 QY 2701 QAKQNVGSGVPMRTVGLNRLTSTFIQVADPDQKGTETKPGQNNPVPVSETNESPIVERT 2760
 Db 2701 QAKQNVGSGVPMRTVGLNRLTSTFIQVADPDQKGTETKPGQNNPVPVSETNESPIVERT 2760
 QY 2761 PFSSSSSKHSSPFGTVAARVTPFNYPNPSRKSSADTSARPSQIPTPVNNTKRDST 2820
 Db 2761 PFSSSSSKHSSPFGTVAARVTPFNYPNPSRKSSADTSARPSQIPTPVNNTKRDST 2820
 QY 2821 DSTESSGTQSPKRGSGYLTVSV 2843
 Db 2821 DSTESSGTQSPKRGSGYLTVSV 2843

RESULT 2
 ID AAW35392

XX AAW35392 standard; protein; 2843 AA.

AC AAW35392;

XX 25-MAR-2003 (revised)

DT 11-MAR-1998 (first entry)

XX Human adenomatous Polyposis coli gene product.

DE Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;

XX neoplastic tissue; tumour tissue; tumour repressor; mutation;

KW sporadic colorectal cancer; detection.

XX Homo sapiens.

OS US5648212-A.

XX

PN

XX 15-JUL-1997.
 Pd 12-AUG-1994; 94US-00289548.
 Pf 16-JAN-1991; 91GB-00000962.
 Pr 16-JAN-1991; 91GB-00000963.
 Pr 16-JAN-1991; 91GB-00000974.
 Pr 16-JAN-1991; 91GB-00000975.
 Pr 08-AUG-1991; 91US-00741940.
 XX (NICA-) JAPANESE FOUND CANCER RES.
 Pa (UTAH) UNIV UTAH.
 Pa (UYJO) UNIV JOHNS HOPKINS.
 Pa (ZENE) ZENECA LTD.
 XX Markham A, Nakamura Y, Groden J, Carlson M, Kinzler K;
 Pi Albertsen H, Hedge PJ, Vogelstein B, Thliveris A, Anand R, White RL;
 Pi Joslyn G;
 XX WFI; 1997-372053/34.
 Dr N-PSDB; AAT95538.
 XX Cancer diagnosis - by detecting mutation(s) in adenomatous polyposis coli gene.
 Pt
 XX Example 1; Col 33-52; 140pp; English.
 Ps The present sequence is the human adenomatous Polyposis coli (APC) gene product, which was used in the development of a novel method of diagnosing or prognosing an APC gene associated neoplastic tissue. The method comprises comparing APC gene coding sequences or mRNA in a tumour tissue, to APC gene coding sequences or mRNA in a non-neoplastic tissue, where a difference indicates an APC gene associated neoplasia of the tumour tissue. APC is a tumour repressor expressed in most normal tissues. APC mutations are found in familial adenomatous polyposis and sporadic colorectal cancer patients. The method enables mutations to be detected to provide an indication of predisposition to cancer. (Updated on 25-MAR-2003 to correct PR field.)
 XX
 Sg Sequence 2843 AA;

Query Match 100.0%; Score 14575; DB 2; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAAASVDOLLKQVEALKWENSLRQELNEDNSNHLTKLETEASNKVEYLKOLGSIIDEAM 60
 Db 1 MAAASVDOLLKQVEALKWENSLRQELNEDNSNHLTKLETEASNKVEYLKOLGSIIDEAM 60
 Qy 61 ASSQIDLLERLKLNLDSNFPFVKLRKMSLRSYSGREGSVSSRSGECSPPVMSGFPR 120
 Db 61 ASSQIDLLERLKLNLDSNFPFVKLRKMSLRSYSGREGSVSSRSGECSPPVMSGFPR 120
 Qy 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKEKDYVAOLNLTKRIDSLPTENFSL 180
 Db 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKEKDYVAOLNLTKRIDSLPTENFSL 180
 Qy 181 QTDLTRRQLEVEARQIRVAMEEQIGTCQDMKEKAQRRIARIQIQIEKDILAIRQLQSQAT 240
 Db 181 QTDLTRRQLEVEARQIRVAMEEQIGTCQDMKEKAQRRIARIQIQIEKDILAIRQLQSQAT 240
 Qy 241 EAERSSQNKHETGSHDAERONEGQGVGEINMATSNGCGSTTRMDHETASVSSSTHSA 300
 Db 241 EAERSSQNKHETGSHDAERONEGQGVGEINMATSNGCGSTTRMDHETASVSSSTHSA 300
 Qy 301 PRLTSHLGTKEVMYVLSLLMLGTHDKDDMSRTLLAMSSQDSCISMROGCLPLLIQLL 360
 Db 301 PRLTSHLGTKEVMYVLSLLMLGTHDKDDMSRTLLAMSSQDSCISMROGCLPLLIQLL 360
 Qy 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHSQPDKRGRRIRVHLLEQIRAYCETC 420
 Db 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHSQPDKRGRRIRVHLLEQIRAYCETC 420

QY	421	WEWQEAHPGMDQDNPM	PAVEHQI	CPAVCVLMKLSFDEBHRHAMNELGGLQIAIELLQ	480
Db	421	WEWQEAHPGMDQDNPM	PAVEHQI	CPAVCVLMKLSFDEBHRHAMNELGGLQIAIELLQ	480
QY	481	VDCEMYGLTNDHYSTILRR	YAGMALTNLT	TFGDVANKATLCSMKGCMRALVAQLKSESEDL	540
Db	481	VDCEMYGLTNDHYSTILRR	YAGMALTNLT	TFGDVANKATLCSMKGCMRALVAQLKSESEDL	540
QY	541	QQVTASVLRLNSWRADVNS	SKTILREVGS	VKALMCALEVKKESTLKSVL\$ALMNL\$AHCT	600
Db	541	QQVTASVLRLNSWRADVNS	SKTILREVGS	VKALMCALEVKKESTLKSVL\$ALMNL\$AHCT	600
QY	601	ENKADICAVDICALAFVGLT	LYRSTQNTLAI	IESGGILRVNSSLIATNEDHROI\$RENN	660
Db	601	ENKADICAVDICALAFVGLT	LYRSTQNTLAI	IESGGILRVNSSLIATNEDHROI\$RENN	660
QY	661	CLQTLLOHLKSH\$TITVSN	ACGTLWNLSARN	PKDOEALWDNGAV\$MLKNLH\$SKH\$MIAM	720
Db	661	CLQTLLOHLKSH\$TITVSN	ACGTLWNLSARN	PKDOEALWDNGAV\$MLKNLH\$SKH\$MIAM	720
QY	721	G\$AAALRLNLMANRP	AKYK\$DANIM\$PG	SGLPSLHV\$KOKALAEADAOHLSETFDNI\$NLS	780
Db	721	G\$AAALRLNLMANRP	AKYK\$DANIM\$PG	SGLPSLHV\$KOKALAEADAOHLSETFDNI\$NLS	780
QY	781	PKASHRSKQRHKQSLYGDY	VFTNRHDDNRSDNF	TNGMTVLSPYLNTTVLPSSSSSRGS	840
Db	781	PKASHRSKQRHKQSLYGDY	VFTNRHDDNRSDNF	TNGMTVLSPYLNTTVLPSSSSSRGS	840
QY	841	LD\$RS\$KDR\$LERGIGL	GNYPATENPCT\$	SKRGLOI\$STTAAQIAKWE\$EVS\$AHTS	900
Db	841	LD\$RS\$KDR\$LERGIGL	GNYPATENPCT\$	SKRGLOI\$STTAAQIAKWE\$EVS\$AHTS	900
QY	901	QEDRS\$G\$STTELHCV	TDERNALR\$SSAAH	TSNTYNTFK\$BNSNRT\$CMPYAKLEYKRS\$	960
Db	901	QEDRS\$G\$STTELHCV	TDERNALR\$SSAAH	TSNTYNTFK\$BNSNRT\$CMPYAKLEYKRS\$	960
QY	961	ND\$LSNVSSNDGYKRG	OMKPSIESY	SEDD\$KFC\$YGOVPADLAHKH\$ANHMDNDGE	1020
Db	961	ND\$LSNVSSNDGYKRG	OMKPSIESY	SEDD\$KFC\$YGOVPADLAHKH\$ANHMDNDGE	1020
QY	1021	LDTPI\$NLSKY\$D\$QL	NSGR\$P\$QNERWAR	PKHII\$EDEIKQ\$BQR\$SRNQ\$STTYPYTE	1080
Db	1021	LDTPI\$NLSKY\$D\$QL	NSGR\$P\$QNERWAR	PKHII\$EDEIKQ\$BQR\$SRNQ\$STTYPYTE	1080
QY	1081	STDDKHLKFQHPGQ	ECV\$PYR\$RGANG	SETNRV\$G\$NHGINQV\$OSLCOEDDYEDDKP	1140
Db	1081	STDDKHLKFQHPGQ	ECV\$PYR\$RGANG	SETNRV\$G\$NHGINQV\$OSLCOEDDYEDDKP	1140
QY	1141	TNYERY\$E\$BQHEE	BERPTNYSKYNEE	KRHVDQPIDY\$LYATDIPSSQKQ\$SF\$SK\$	1200
Db	1141	TNYERY\$E\$BQHEE	BERPTNYSKYNEE	KRHVDQPIDY\$LYATDIPSSQKQ\$SF\$SK\$	1200
QY	1201	SSG\$S\$KTEH\$SSSEN	STP\$SSNAKQNLH	PP\$SAQ\$SRGQ\$POKATCKV\$SNQB\$TIQ	1260
Db	1201	SSG\$S\$KTEH\$SSSEN	STP\$SSNAKQNLH	PP\$SAQ\$SRGQ\$POKATCKV\$SNQB\$TIQ	1260
QY	1261	TYCVEDTPICF\$SR	CS\$LSL\$S\$A\$DE	IGC\$NQTQ\$EAD\$SANTLOI\$AEIKGIGTR\$AEDPV	1320
Db	1261	TYCVEDTPICF\$SR	CS\$LSL\$S\$A\$DE	IGC\$NQTQ\$EAD\$SANTLOI\$AEIKGIGTR\$AEDPV	1320
QY	1321	SEVP\$AV\$QHPRTK\$SR	LOG\$S\$LS\$ES	ARHKA\$VFP\$PSGAK\$SP\$SK\$GAOTPK\$PEHYVQET	1380
Db	1321	SEVP\$AV\$QHPRTK\$SR	LOG\$S\$LS\$ES	ARHKA\$VFP\$PSGAK\$SP\$SK\$GAOTPK\$PEHYVQET	1380
QY	1381	PLM\$P\$RCT\$Y\$SS	LDS\$F\$SR\$IAS	SQ\$EPC\$G\$MV\$SGII\$S\$DLP\$D\$SQ\$TMP\$SR\$SKTTP	1440
Db	1381	PLM\$P\$RCT\$Y\$SS	LDS\$F\$SR\$IAS	SQ\$EPC\$G\$MV\$SGII\$S\$DLP\$D\$SQ\$TMP\$SR\$SKTTP	1440
QY	1441	PPQ\$TAOTK\$EVPK	AKPTAEK	R\$E\$G\$PKQ\$AVNA\$VORVQL\$PDADTLLHFA\$TESTPDGF	1500
Db	1441	PPQ\$TAOTK\$EVPK	AKPTAEK	R\$E\$G\$PKQ\$AVNA\$VORVQL\$PDADTLLHFA\$TESTPDGF	1500

QY	1501	SCSSLSALS	LDPEFIQKV	VELRIMPPVQ	ENDNGNETSE	SOPKES	ENQKEA	KITDSE	1560																																	
QY	1501	SCSSLSALS	LDPEFIQKV	VELRIMPPVQ	ENDNGNETSE	SOPKES	ENQKEA	KITDSE	1560																																	
QY	1561	KDLDDSD	DDDDIE	LEBCEIIS	AMPTKSS	RKGKPAQ	ATASKLPP	PPVARKPSQ	LPVYKLLPS	1620																																
QY	1561	KDLDDSD	DDDDIE	LEBCEIIS	AMPTKSS	RKGKPAQ	ATASKLPP	PPVARKPSQ	LPVYKLLPS	1620																																
QY	1621	QNRLOPQ	KVSVFT	PGDDMP	RVYCV	VEGPIIN	FSTATS	LSDLTIB	SSPNELAA	GCVRGGAQ	1680																															
QY	1621	QNRLOPQ	KVSVFT	PGDDMP	RVYCV	VEGPIIN	FSTATS	LSDLTIB	SSPNELAA	GCVRGGAQ	1680																															
QY	1681	SGEFEK	DDTTP	TEGRST	DEAOG	GKTSSV	TLPEL	DDNKABEG	DILAE	CINISAM	PKGKSHKP	1740																														
QY	1681	SGEFEK	DDTTP	TEGRST	DEAOG	GKTSSV	TLPEL	DDNKABEG	DILAE	CINISAM	PKGKSHKP	1740																														
QY	1741	FRVKKIM	DVQOAS	ASSAP	NKNQ	LDGKKK	PTSPVK	PIPONTE	YRTRV	KEKNAD	SNNLN	1800																														
QY	1741	FRVKKIM	DVQOAS	ASSAP	NKNQ	LDGKKK	PTSPVK	PIPONTE	YRTRV	KEKNAD	SNNLN	1800																														
QY	1801	AERVFS	DNKSKQ	NLKN	NSKOF	NDKLP	NNED	VRGS	FAPDS	PHHTPT	EGT	PYCHSRND	1860																													
QY	1801	AERVFS	DNKSKQ	NLKN	NSKOF	NDKLP	NNED	VRGS	FAPDS	PHHTPT	EGT	PYCHSRND	1860																													
QY	1861	SLSLDP	DDDDV	DL	SREK	ABLRA	KENKE	SEAKVT	SHTELTS	NOQANK	TOAIA	KOPINR	1920																													
QY	1861	SLSLDP	DDDDV	DL	SREK	ABLRA	KENKE	SEAKVT	SHTELTS	NOQANK	TOAIA	KOPINR	1920																													
QY	1921	GQPKPIL	QKOST	TPQSS	KDIP	DRC	GAAT	DEKLON	PAIENT	PVCF	SHNSS	LSLSDID	QENN	1980																												
QY	1921	GQPKPIL	QKOST	TPQSS	KDIP	DRC	GAAT	DEKLON	PAIENT	PVCF	SHNSS	LSLSDID	QENN	1980																												
QY	1981	NKENEP	I	KETEP	PD	SQGE	PSKPA	QASG	VAPKS	FVEVD	TPVCF	SNSLS	LSLSDID	LLQ	2040																											
QY	1981	NKENEP	I	KETEP	PD	SQGE	PSKPA	QASG	VAPKS	FVEVD	TPVCF	SNSLS	LSLSDID	LLQ	2040																											
QY	2041	ECISS	AMPKK	KKPSRL	KDNE	KHS	PRNM	GILGED	LTLD	KD	IORD	PDE	SEHGL	SPDSENFD	2100																											
QY	2041	ECISS	AMPKK	KKPSRL	KDNE	KHS	PRNM	GILGED	LTLD	KD	IORD	PDE	SEHGL	SPDSENFD	2100																											
QY	2101	WKAIQ	EGANS	I	VSLHQ	AAAA	CL	SQAS	SDS	ILSL	XSGIS	SLG	SPFHL	TPD	KEKPT	2160																										
QY	2101	WKAIQ	EGANS	I	VSLHQ	AAAA	CL	SQAS	SDS	ILSL	XSGIS	SLG	SPFHL	TPD	KEKPT	2160																										
QY	2161	SNKGPR	ILX	PGEX	STL	ETK	IE	SESG	XG	IKG	KVYK	SLIT	GVR	NS	ISG	WKQPLOAN	2220																									
QY	2161	SNKGPR	ILX	PGEX	STL	ETK	IE	SESG	XG	IKG	KVYK	SLIT	GVR	NS	ISG	WKQPLOAN	2220																									
QY	2221	MPGIS	RGR	THI	H	PGV	RNS	SSST	PSVSK	GPPLK	TP	P	AKSP	SB	QAT	TATTS	PRGAK	PSVKS	2280																							
QY	2221	MPGIS	RGR	THI	H	PGV	RNS	SSST	PSVSK	GPPLK	TP	P	AKSP	SB	QAT	TATTS	PRGAK	PSVKS	2280																							
QY	2281	ELSPV	ARQ	TSO	IG	SSK	AP	SR	SG	RST	PSR	PA	QPL	SR	I	Q	SP	GR	NS	I	SP	GR	NG	I	SP	PN	2340															
QY	2281	ELSPV	ARQ	TSO	IG	SSK	AP	SR	SG	RST	PSR	PA	QPL	SR	I	Q	SP	GR	NS	I	SP	GR	NG	I	SP	PN	2340															
QY	2341	KL	SQL	PR	T	SP	ST	AST	K	SG	SG	MY	T	SP	Q	R	MS	Q	N	L	T	K	Q	T	G	L	S	K	V	AS	I	PR	SE	S	A	S	K	G	2400			
QY	2341	KL	SQL	PR	T	SP	ST	AST	K	SG	SG	MY	T	SP	Q	R	MS	Q	N	L	T	K	Q	T	G	L	S	K	V	AS	I	PR	SE	S	A	S	K	G	2400			
QY	2401	LNQMN	NG	N	G	N	K	K	VEL	SR	MS	ST	SG	S	E	S	D	R	S	R	P	V	L	R	Q	S	T	I	K	E	A	P	T	L	R	K	L	E	E	S	A	2460

Db 2581 SERAKSEDEKRVNSISGTQSKENQVSAKGTWRKIKENEFSPTNSTQTVSSGATNGAES 2640
Qy 2641 KTLIIQVAPAVSKTDEWVRIEDCPINNPRSGRSPGNTPTPPVIDSVSEKANPNIKDSKDN 2700
Db 2641 KTLIIQVAPAVSKTDEWVRIEDCPINNPRSGRSPGNTPTPPVIDSVSEKANPNIKDSKDN 2700
Qy 2701 QAKQNVGSGVPMRTVGLNRLTSFIQVDAPDQKGTBIKPGQNNPVPVSTNESPVERT 2760
Db 2701 QAKQNVGSGVPMRTVGLNRLTSFIQVDAPDQKGTBIKPGQNNPVPVSTNESPVERT 2760
Qy 2761 PFSSSSSKHSSPSGTVAAARVTPNNYKPSPRKSSADSTSARPSQIPTPVNNTKKRDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAAARVTPNNYKPSPRKSSADSTSARPSQIPTPVNNTKKRDSKT 2820
Qy 2821 DSTESSGTQPKRHSGSYLTVSV 2843
Db 2821 DSTESSGTQPKRHSGSYLTVSV 2843
RESULT 3
AAW38370
ID AAW38370 standard; protein; 2843 AA.
XX
AC AAW38370;
DT 25-MAR-2003 (revised)
DT 08-APR-1998 (first entry)
XX
DE Human adenomatous Polyposis coli gene product.
XX
KW Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;
KW neoplastic tissue; tumour tissue; tumour repressor; mutation;
KW sporadic colorectal cancer; detection.
XX
OS Homo sapiens.
XX
PN US5691454-A.
XX
PD 25-NOV-1997.
XX
PF 25-MAY-1995; 95US-00452854.
XX
PR 16-JAN-1991; 91GB-00000962.
PR 16-JAN-1991; 91GB-00000963.
PR 16-JAN-1991; 91GB-00000974.
PR 16-JAN-1991; 91GB-00000975.
PR 08-AUG-1991; 91US-00741840.
PR 12-AUG-1994; 94US-00289548.
XX
PA (CANC-) CANCER INST.
PA (UYJO) UNIV JOHNS HOPKINS.
PA (ICIL) IMPERIAL CHEM IND PLC.
PA (UTAH) UNIV UTAH.
XX
Pi Nakamura Y, Markham AP, Groden J, Vogelstein B, Kinzler K,
Pi Hedge PJ, Carlson M, Anand R, Thliveris A, Albertsen H, White RL;
Pi Joselyn G;
XX
DR WPI; 1998-017712/02.
DR N-FSDB; AAT96153.
XX
PT Antibodies to normal and mutant adenomatous polyposis coli proteins -
PT useful for detecting genetic predisposition to cancer.
XX
PS Example 1; Col 33-52; 107pp; English.
XX
CC The present sequence is the human adenomatous Polyposis coli (APC) gene
CC product, which was used in the development of a novel method of
CC diagnosing or prognosing an APC gene associated neoplastic tissue. The
CC method comprises comparing APC gene coding sequences or mRNA in a tumour
CC tissue, to APC gene coding sequences or mRNA in a non-neoplastic tissue,
CC where a difference indicates an APC gene associated neoplasia of the

CC tumour tissue. APC is a tumour repressor expressed in most normal
CC tissues. APC mutations are found in familial adenomatous polyposis and
CC sporadic colorectal cancer patients. The method enables mutations to be
CC detected to provide an indication of predisposition to cancer. (Updated
CC on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2843 AA;
Query Match 100.0%; Score 14575; DB 2; Length 2843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAASVDOLLKQVEALKWENSLROLEDNSNHLTKLEASNMKEVLKQLGSGTDEAM 60
Db 1 MAASVDOLLKQVEALKWENSLROLEDNSNHLTKLEASNMKEVLKQLGSGTDEAM 60
Qy 61 ASSGQIDLLERLKEINLDSSNPPGPKLRSKMSLRSYGSRGSSVSSRSGECPVPMGSPFR 120
Db 61 ASSGQIDLLERLKEINLDSSNPPGPKLRSKMSLRSYGSRGSSVSSRSGECPVPMGSPFR 120
Qy 121 RGFVNGSRSTCYLEELKERSLLADLDLKEEKEKDWYQAQNLTKRIDSILPTENSL 180
Db 121 RGFVNGSRSTCYLEELKERSLLADLDLKEEKEKDWYQAQNLTKRIDSILPTENSL 180
Qy 181 QTDLTRQLEYEARQIRVAMEBOLGTCODMEKRAQRRIARIQIEKDILRIRQLLSQAT 240
Db 181 QTDLTRQLEYEARQIRVAMEBOLGTCODMEKRAQRRIARIQIEKDILRIRQLLSQAT 240
Qy 241 EAERSQNKHETGSHDAERQNEGQGVGEINMATSGNGQSTTRMDHETASVLSSTSTSA 300
Db 241 EAERSQNKHETGSHDAERQNEGQGVGEINMATSGNGQSTTRMDHETASVLSSTSTSA 300
Qy 301 PRRLTSHLGTKEVMYVYLLSMGLTHDKDMSRTLLAMSSQDSCISMRQSGCLPLIQLL 360
Db 301 PRRLTSHLGTKEVMYVYLLSMGLTHDKDMSRTLLAMSSQDSCISMRQSGCLPLIQLL 360
Qy 361 HGNDKDSVLLGNSRGSKEARAPASAAALHNI IHSQDDKRGREIRVLLHLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNSRGSKEARAPASAAALHNI IHSQDDKRGREIRVLLHLEQIRAYCETC 420
Qy 421 WEQBAHEPGMDQDNPMAPVVEHQICPAVCVLMKLSFDEEHRHAMELGGHQAELLO 480
Db 421 WEQBAHEPGMDQDNPMAPVVEHQICPAVCVLMKLSFDEEHRHAMELGGHQAELLO 480
Qy 481 VDCEMYGLTNDHYSITLRYAGVALTNTLTFGVANKATLCSMKGMALVAQLKSESD 540
Db 481 VDCEMYGLTNDHYSITLRYAGVALTNTLTFGVANKATLCSMKGMALVAQLKSESD 540
Qy 541 QQVIASVLENLNWRADVNSKTLREVGSVKALMECALEVKKESTLKSVLNLSAHL 600
Db 541 QQVIASVLENLNWRADVNSKTLREVGSVKALMECALEVKKESTLKSVLNLSAHL 600
Qy 601 ENKADICAVDGLAFVGLTYRSOTNTLAIIESGGILRNVSSTIATNEDHRQLRENN 660
Db 601 ENKADICAVDGLAFVGLTYRSOTNTLAIIESGGILRNVSSTIATNEDHRQLRENN 660
Qy 661 CLQTLQHLKSHSLTIVSNACOTLWNLNARPNKQDEALWDMGAVSNLKNL IHSKHMTAM 720
Db 661 CLQTLQHLKSHSLTIVSNACOTLWNLNARPNKQDEALWDMGAVSNLKNL IHSKHMTAM 720
Qy 721 GSAALRNLMANRPAYKDANIMSPGSLPSLHVKKOKALEBAELDAQHLSETFDINDLS 780
Db 721 GSAALRNLMANRPAYKDANIMSPGSLPSLHVKKOKALEBAELDAQHLSETFDINDLS 780
Qy 781 PKASHRSKORHKQSLYGDYVFTDNRHDDNRSDFNTGNMTVLSPLYNTTLPSSSSSRGS 840
Db 781 PKASHRSKORHKQSLYGDYVFTDNRHDDNRSDFNTGNMTVLSPLYNTTLPSSSSSRGS 840
Qy 841 LSSSRSEKDRSLEREREGIGLGNHYHATENPGTSSKRGLOISTAAQIAKMEVSAIHTS 900
Db 841 LSSSRSEKDRSLEREREGIGLGNHYHATENPGTSSKRGLOISTAAQIAKMEVSAIHTS 900
Qy 901 QEDRSSGSGTTELHCVTDERNALRRSSAAHTSNTYNTFKSENSNRTCSMPYAKLEYKRSS 960

Db 901 QEDRSSGSGTTELHCVTDERNALRRSAAHTHNTYNTKSENSTCSMPYAKLYKRS 960
 Qy 961 NDSLSVSSNDGCGKQGMKPSIESYSEDESKFCSYGOYPADLAHKIHSANMMDNDGE 1020
 Db 961 NDSLSVSSNDGCGKQGMKPSIESYSEDESKFCSYGOYPADLAHKIHSANMMDNDGE 1020
 Qy 1021 LDTPIYSLKYSEQLNSGRQSPSONERWARPKHIIEDIKOSEORQSNQSTTVPVYTE 1080
 Db 1021 LDTPIYSLKYSEQLNSGRQSPSONERWARPKHIIEDIKOSEORQSNQSTTVPVYTE 1080
 Qy 1081 STDKHLKFOHPHQOECVSPYRSRGANGSETRVGSNGINONVSQSLCQEDDVEDDKP 1140
 Db 1081 STDKHLKFOHPHQOECVSPYRSRGANGSETRVGSNGINONVSQSLCQEDDVEDDKP 1140
 Qy 1141 TNYSEYSEEEQHEBERPTNYSIKYNEEKRVDPIDYSLKYATDIPSSQKOSFSFKS 1200
 Db 1141 TNYSEYSEEEQHEBERPTNYSIKYNEEKRVDPIDYSLKYATDIPSSQKOSFSFKS 1200
 Qy 1201 SSGOSKTHMSSSENTSTPSSNAKRONLHPSSAQSRSGOPKAATCKUSSINOETIQ 1260
 Db 1201 SSGOSKTHMSSSENTSTPSSNAKRONLHPSSAQSRSGOPKAATCKUSSINOETIQ 1260
 Qy 1261 TYCVEDTPICFGRCSLSLSAEDIGCNQTTQADSAANTIQAIEIKIGKIGTRSAEDPV 1320
 Db 1261 TYCVEDTPICFGRCSLSLSAEDIGCNQTTQADSAANTIQAIEIKIGKIGTRSAEDPV 1320
 Qy 1321 SEVPAYSQHPRTKSSRLQSSLSSESARHKAFFPSGAKSPKSGAOTPKSPPEHYVQET 1380
 Db 1321 SEVPAYSQHPRTKSSRLQSSLSSESARHKAFFPSGAKSPKSGAOTPKSPPEHYVQET 1380
 Qy 1381 PLMFSSCTSVSSLDSPESRSIASSVQSPGSGVSGIISPSDLPSDQTMPPSRSKTTP 1440
 Db 1381 PLMFSSCTSVSSLDSPESRSIASSVQSPGSGVSGIISPSDLPSDQTMPPSRSKTTP 1440
 Qy 1441 PPQTAQTRVPKNAKPTAEKRESGPKQAANAAVQRVQVLPDADTLHLFATETPDGF 1500
 Db 1441 PPQTAQTRVPKNAKPTAEKRESGPKQAANAAVQRVQVLPDADTLHLFATETPDGF 1500
 Qy 1501 SCSSLSALSLEDPFQKQVELRIMPVQENDNGNETSEQPKENQEKAEKTIIDSE 1560
 Db 1501 SCSSLSALSLEDPFQKQVELRIMPVQENDNGNETSEQPKENQEKAEKTIIDSE 1560
 Qy 1561 KDLDDSDDDDIIEBECIISAMPTKSRGKPKAQTASKLPPVARKPSQLPVYKLLPS 1620
 Db 1561 KDLDDSDDDDIIEBECIISAMPTKSRGKPKAQTASKLPPVARKPSQLPVYKLLPS 1620
 Qy 1621 QNLPQKQVSTPFGDMPRVYCVETGPIINFSTATSLDLTIESPPNELAAGEGVRGGAQ 1680
 Db 1621 QNLPQKQVSTPFGDMPRVYCVETGPIINFSTATSLDLTIESPPNELAAGEGVRGGAQ 1680
 Qy 1681 SGFEKRDITPTEGRSTDEAQGKTSSTVPIBELDKNKABEGDILAEICINSAMPKGSHP 1740
 Db 1681 SGFEKRDITPTEGRSTDEAQGKTSSTVPIBELDKNKABEGDILAEICINSAMPKGSHP 1740
 Qy 1741 FRVKKIMDQVQASASSAPNKNQLDGKKKPTSPVKPIPQNTYRTRVRKNADSNNLN 1800
 Db 1741 FRVKKIMDQVQASASSAPNKNQLDGKKKPTSPVKPIPQNTYRTRVRKNADSNNLN 1800
 Qy 1801 AERVFSDNKKQNLKNSKDFNDKLPNNEDRVGSAFDPSPHYTPIEGTPYCFSRND 1860
 Db 1801 AERVFSDNKKQNLKNSKDFNDKLPNNEDRVGSAFDPSPHYTPIEGTPYCFSRND 1860
 Qy 1861 SLSSLDFFDDDDVLSREKAELEKAKENKSEAKVTSHTBELTSNQOSANKTOIAKQPINR 1920
 Db 1861 SLSSLDFFDDDDVLSREKAELEKAKENKSEAKVTSHTBELTSNQOSANKTOIAKQPINR 1920
 Qy 1921 GQPKPILOKQSTFPQSSKDIIPRGAATDEKLNFAIENTPVCFSNSSLSSLDIDQENN 1980
 Db 1921 GQPKPILOKQSTFPQSSKDIIPRGAATDEKLNFAIENTPVCFSNSSLSSLDIDQENN 1980
 Qy 1981 NKNEPIKETEPPDPSQGEPSKQASGYAPKSFHVBDTPVCFPSRNSLSLSDSEDDLLQ 2040

Db 1981 NKNEPIKETEPPDPSQGEPSKQASGYAPKSFHVBDTPVCFPSRNSLSLSDSEDDLLQ 2040
 Qy 2041 ECISAMPKKKPSRLKGDNEKHSPRNMGILGEDLTLDLKDQIRPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKKKPSRLKGDNEKHSPRNMGILGEDLTLDLKDQIRPDSEHGLSPDSENF 2100
 Qy 2101 WKAIQEGANSIVSLHQAAAAACLSRQASDSDSILSLKSGISLSPFHLTPDQEKXPT 2160
 Db 2101 WKAIQEGANSIVSLHQAAAAACLSRQASDSDSILSLKSGISLSPFHLTPDQEKXPT 2160
 Qy 2161 SNKGPRLLKPEGEKSTLETKKIESEKIGKGGKVKYKSLITGKVRNSEISGQMKQPLQAN 2220
 Db 2161 SNKGPRLLKPEGEKSTLETKKIESEKIGKGGKVKYKSLITGKVRNSEISGQMKQPLQAN 2220
 Qy 2221 MPSISRGRTMIHIPGVNRNSSSTSPVSKGPPPLKTPASKSPSEGOATTSPRGAKEPVKS 2280
 Db 2221 MPSISRGRTMIHIPGVNRNSSSTSPVSKGPPPLKTPASKSPSEGOATTSPRGAKEPVKS 2280
 Qy 2281 ELSFVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPIOSFGRNSISPGRNGISPPN 2340
 Db 2281 ELSFVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPIOSFGRNSISPGRNGISPPN 2340
 Qy 2341 KLSQLPRTSPSTASTKSGSGKVSYSYTSRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
 Db 2341 KLSQLPRTSPSTASTKSGSGKVSYSYTSRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
 Qy 2401 LNMQNGNGANKKVELSRMSSTKSGSGESDRSEPVLRQSTFIKEAPSTLRKLEESA 2460
 Db 2401 LNMQNGNGANKKVELSRMSSTKSGSGESDRSEPVLRQSTFIKEAPSTLRKLEESA 2460
 Qy 2461 SFESLSPSSPASPTRSAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Db 2461 SFESLSPSSPASPTRSAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Qy 2521 RPAKHDIASHSESRLPINSRGTWKREHSHSSSLPRVSTWRTGSSSILSASSES 2580
 Db 2521 RPAKHDIASHSESRLPINSRGTWKREHSHSSSLPRVSTWRTGSSSILSASSES 2580
 Qy 2581 SERAKSEDEKHVANSISGTSKQENQVSAKTWTKIKENEFSPNTNSTQVSSGATNGAES 2640
 Db 2581 SERAKSEDEKHVANSISGTSKQENQVSAKTWTKIKENEFSPNTNSTQVSSGATNGAES 2640
 Qy 2641 KTLIYQAPAVSKTEDVWVRIEDCPINNPRSGSPGTNTPPVDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIYQAPAVSKTEDVWVRIEDCPINNPRSGSPGTNTPPVDSVSEKANPNIKDSKN 2700
 Qy 2701 QAKQNVGNGVPMRTVGLNRLTSFIQVDAPDQKGTETKPGQNNPVVSETNESPIVERT 2760
 Db 2701 QAKQNVGNGVPMRTVGLNRLTSFIQVDAPDQKGTETKPGQNNPVVSETNESPIVERT 2760
 Qy 2761 PFSSSSSKHSSPSGTVAARVTPFNPNPRKSSADSTARSQIPTPVNNNTKXDSKT 2820
 Db 2761 PFSSSSSKHSSPSGTVAARVTPFNPNPRKSSADSTARSQIPTPVNNNTKXDSKT 2820
 Qy 2821 DSTESSGTQSPKRSGLYVTSV 2843
 Db 2821 DSTESSGTQSPKRSGLYVTSV 2843
 RESULT 4
 ABG90964
 ID ABG90964 standard; protein; 2843 AA.
 XX AC ABG90964;
 XX XX
 DT 29-NOV-2002 (first entry)
 XX XX
 DE Human adenomatous polyposis coli (APC) protein #1.
 XX KW Adenomatous polyposis coli; APC; human; neoplastic tissue;
 XX KW mutation detection; tumour; cancer.
 OS Homo sapiens.

DB	361	HNENDKXSVLLGNRSRGSKEAPARASAAALHNI	THSQPDDKRGREIRVLLHLLQIRAYCETC	420			
QY	421	NEWQBAEPFGMDQDKNPPAPVEHQICPAVCVLMKLS	SPDEBHRHAMNELGGIQAIAELLQ	480			
DB	421	NEWQBAEPFGMDQDKNPPAPVEHQICPAVCVLMKLS	SPDEBHRHAMNELGGIQAIAELLQ	480			
QY	481	VDCEMYGILTNDRHYSITLIRYAGMALTNLT	TFGDVANAKATLCMKGCGRALVAQLKSESEDL	540			
DB	481	VDCEMYGILTNDRHYSITLIRYAGMALTNLT	TFGDVANAKATLCMKGCGRALVAQLKSESEDL	540			
QY	541	QOVIASVLRLNLSWRADVNSKKT	LRVGSVKALMECALEVKKESTLKSVL	SALWNL	SAHCT	600	
DB	541	QOVIASVLRLNLSWRADVNSKKT	LRVGSVKALMECALEVKKESTLKSVL	SALWNL	SAHCT	600	
QY	601	ENKADI	CAVDGALAPLVGTLTYRSTQNTLAI	IESGGGILRNVS	LIATNE	DHRQILRENN	660
DB	601	ENKADI	CAVDGALAPLVGTLTYRSTQNTLAI	IESGGGILRNVS	LIATNE	DHRQILRENN	660
QY	661	CLQOTLLOHLKHSLSLTVSNACGTL	NWLSARNPQOEALWDMGAVSMKLN	LHSHKHMTAM			720
DB	661	CLQOTLLOHLKHSLSLTVSNACGTL	NWLSARNPQOEALWDMGAVSMKLN	LHSHKHMTAM			720
QY	721	GSAAALRLMANRPAKYKDANIMSGSSLP	PSLHVKKQKALAEALDAQHLSETFD	NIDNLS			780
DB	721	GSAAALRLMANRPAKYKDANIMSGSSLP	PSLHVKKQKALAEALDAQHLSETFD	NIDNLS			780
QY	781	PKASHRSKORHKOSLYGDIYVDFTN	RHDNRSDNFNTGNMTVLS	PYLNTT	TVLPSSSS	SRGS	840
DB	781	PKASHRSKORHKOSLYGDIYVDFTN	RHDNRSDNFNTGNMTVLS	PYLNTT	TVLPSSSS	SRGS	840
QY	841	LDSRSSEKDRSLERERGI	GLGNYHPATENPGTSSKRG	LQISTTAAQIAK	MVEE	SAIHTS	900
DB	841	LDSRSSEKDRSLERERGI	GLGNYHPATENPGTSSKRG	LQISTTAAQIAK	MVEE	SAIHTS	900
QY	901	QDRSGSGTTEHLCVTDERNALRR	SSAAHTSNNTYNFTKSENSNRT	CMPIAK	LEYK	SS	960
DB	901	QDRSGSGTTEHLCVTDERNALRR	SSAAHTSNNTYNFTKSENSNRT	CMPIAK	LEYK	SS	960
QY	961	NDSLNSVSNDOGYKRGQWKPSI	IESYSDDESKFC	SYGOYPADLAH	LKTHSANHMD	NDGE	1020
DB	961	NDSLNSVSNDOGYKRGQWKPSI	IESYSDDESKFC	SYGOYPADLAH	LKTHSANHMD	NDGE	1020
QY	1021	LDTPINYS	LKYSDEQLNSGRSPQNER	WARPKHII	IEDEIKQSEQR	OSRNO	1080
DB	1021	LDTPINYS	LKYSDEQLNSGRSPQNER	WARPKHII	IEDEIKQSEQR	OSRNO	1080
QY	1081	STDDKHLKFPQHFGQOECVSP	VRSGANGSETRVGSNHGINQNV	SQSLCQEDD	YEDDKP		1140
DB	1081	STDDKHLKFPQHFGQOECVSP	VRSGANGSETRVGSNHGINQNV	SQSLCQEDD	YEDDKP		1140
QY	1141	TNYSERYEEEH	EEERPTNYSIKYNEEK	RHVDQPIDYSLKYATD	IPSSQK	SFSGS	1200
DB	1141	TNYSERYEEEH	EEERPTNYSIKYNEEK	RHVDQPIDYSLKYATD	IPSSQK	SFSGS	1200
QY	1201	SSGQSKTEH	MSSENSTPSSNAKRONOLHP	SSAQRSGGPOKAA	CKVSSINQ	ETIQ	1260
DB	1201	SSGQSKTEH	MSSENSTPSSNAKRONOLHP	SSAQRSGGPOKAA	CKVSSINQ	ETIQ	1260
QY	1261	TYCVEDTPTIC	SRCSLSSLSAEDEIGC	NOTTQ	EAD	SANTLIQIAEIKG	1320
DB	1261	TYCVEDTPTIC	SRCSLSSLSAEDEIGC	NOTTQ	EAD	SANTLIQIAEIKG	1320
QY	1321	SEVP	AVSOHPRTKSRLOGSSLS	SESARHKAVEFP	SGAKSP	SGAQT	1380
DB	1321	SEVP	AVSOHPRTKSRLOGSSLS	SESARHKAVEFP	SGAKSP	SGAQT	1380
QY	1381	PLMF	SRCCTSVS	LDSEFERSI	TASSVQSEPC	SGWSGIIT	1440
DB	1381	PLMF	SRCCTSVS	LDSEFERSI	TASSVQSEPC	SGWSGIIT	1440
QY	1441	PPPQ	TAQTKREVPKNKAPTAEK	RESGPKQAAVNA	AVQRVQVLP	DADTLLH	1500
DB	1441	PPPQ	TAQTKREVPKNKAPTAEK	RESGPKQAAVNA	AVQRVQVLP	DADTLLH	1500

Db 1441 PPQTAQTRKVPKNAKPTAEKRESGPKQAAYNAVAQVQVLPDADTLILHPATESTPDGF 1500
QY 1501 SCSSLSALSLDEPFTQKDVLRIMPPVOENDNGNETSEQPKESNENQEKAEKTIIDSE 1560
Db 1501 SCSSLSALSLDEPFTQKDVLRIMPPVOENDNGNETSEQPKESNENQEKAEKTIIDSE 1560
QY 1561 KDLDDSDDDDEILIEECIIISAMPTKSRKPKKPACTAKLPPVARKPSQLPVYKLLPS 1620
Db 1561 KDLDDSDDDDEILIEECIIISAMPTKSRKPKKPACTAKLPPVARKPSQLPVYKLLPS 1620
QY 1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGEGVGGGAQ 1680
Db 1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGEGVGGGAQ 1680
QY 1681 SGFEFKRDTIPTEGRSTDEAQQGKTSSTVPIELDNDKABEGDILAEICINSAMPKSKHP 1740
Db 1681 SGFEFKRDTIPTEGRSTDEAQQGKTSSTVPIELDNDKABEGDILAEICINSAMPKSKHP 1740
QY 1741 FRVKIIMDQVQOASASSAPKNQDGGKKTSPVKPIQNTYRTRVRKXADSKNNLN 1800
Db 1741 FRVKIIMDQVQOASASSAPKNQDGGKKTSPVKPIQNTYRTRVRKXADSKNNLN 1800
QY 1801 AERVFSDNDSKQNLKNNKDFNDKLPNNEDRVGSAFDPSPHYHTPIEGTPYCFSRND 1860
Db 1801 AERVFSDNDSKQNLKNNKDFNDKLPNNEDRVGSAFDPSPHYHTPIEGTPYCFSRND 1860
QY 1861 SLSSLDFFDDDDVLSREKAEELKAKENKSEAKVTSHTLTSNQOSANKTOIAKOPINR 1920
Db 1861 SLSSLDFFDDDDVLSREKAEELKAKENKSEAKVTSHTLTSNQOSANKTOIAKOPINR 1920
QY 1921 GQPKPILOKQSTPQSSKDIIPRGAATDEKLNFAIENTPVCFSHNSLSLSDIDQENN 1980
Db 1921 GQPKPILOKQSTPQSSKDIIPRGAATDEKLNFAIENTPVCFSHNSLSLSDIDQENN 1980
QY 1981 KNEPEIKETEPDSDGEGSKQAAGYAPKSFHVEDTPVCFERNLSLSLSDIDEDLLQ 2040
Db 1981 KNEPEIKETEPDSDGEGSKQAAGYAPKSFHVEDTPVCFERNLSLSLSDIDEDLLQ 2040
QY 2041 ECISSAMPKKKPSRLKGNKEHSPRNMGGIIGEDLTLDKDIORPDSEHGLSPDSENF 2100
Db 2041 ECISSAMPKKKPSRLKGNKEHSPRNMGGIIGEDLTLDKDIORPDSEHGLSPDSENF 2100
QY 2101 WKAIQEGANSIVSSLHQAACLSQASDSDSLKSLGSLGSPHLPDQEKPT 2160
Db 2101 WKAIQEGANSIVSSLHQAACLSQASDSDSLKSLGSLGSPHLPDQEKPT 2160
QY 2161 SNKGPIILPGEKSTLTETKIESESGIKGGKVKYSLITGVRNSLSLSDIDEDLLQ 2220
Db 2161 SNKGPIILPGEKSTLTETKIESESGIKGGKVKYSLITGVRNSLSLSDIDEDLLQ 2220
QY 2221 MPSISGRTHIHPGVNRSSTSPVSKGPKLTPASKSPSEGOTATTSRGAKPSVKS 2280
Db 2221 MPSISGRTHIHPGVNRSSTSPVSKGPKLTPASKSPSEGOTATTSRGAKPSVKS 2280
QY 2281 ELSPVARQTSOIGGSKAPSRGSRDSTPSRAQOPLSRPIQSPGRNISIPGRNISPPN 2340
Db 2281 ELSPVARQTSOIGGSKAPSRGSRDSTPSRAQOPLSRPIQSPGRNISIPGRNISPPN 2340
QY 2341 KLSQLPRTSPSPASTKSSGSGKMSYTPSGRQVSOQNLTKOTGLSKNASSIPRSESASG 2400
Db 2341 KLSQLPRTSPSPASTKSSGSGKMSYTPSGRQVSOQNLTKOTGLSKNASSIPRSESASG 2400
QY 2401 LNMWNGANGKVKELSRMSSSTKSSGSEDRERPVLRQSTFIKEAPSPTRLRKLESA 2460
Db 2401 LNMWNGANGKVKELSRMSSSTKSSGSEDRERPVLRQSTFIKEAPSPTRLRKLESA 2460
QY 2461 SFESLSPSRPAPTRSOAQTPVLSPLDMSLSTHSSVQAGWRKLPNLSPPTIEYNDG 2520
Db 2461 SFESLSPSRPAPTRSOAQTPVLSPLDMSLSTHSSVQAGWRKLPNLSPPTIEYNDG 2520
QY 2521 RPAKRDHIDARSHESPSRLPINPSCGTWKEHSGHSSLPVSTWERTCSSSILSASES 2580
Db 2521 RPAKRDHIDARSHESPSRLPINPSCGTWKEHSGHSSLPVSTWERTCSSSILSASES 2580

QY 2581 SEKAKEDEKHVNSISGTSKOSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
Db 2581 SEKAKEDEKHVNSISGTSKOSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
QY 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
Db 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
QY 2701 QAKQNVGVGSPVMTVTGLENRLTSFTQVDAPDQKGTIKPGQNNPVVSETNESPIVET 2760
Db 2701 QAKQNVGVGSPVMTVTGLENRLTSFTQVDAPDQKGTIKPGQNNPVVSETNESPIVET 2760
QY 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPSRKSSADSTGARSQIPTPVNNNTKKRDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPSRKSSADSTGARSQIPTPVNNNTKKRDSKT 2820
QY 2821 DSTESSGTQSPKHSGLVTSV 2843
Db 2821 DSTESSGTQSPKHSGLVTSV 2843

RESULT 5

AAR63507
ID AAR63507 standard; protein; 2860 AA.

XX AAR63507;

AC AC

DT 25-MAR-2003 (revised)

DT 30-MAY-1995 (first entry)

XX XX

DE Adenomatous polyposis coli tumour repressor.

XX Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;
KW familial adenomatous polyposis; cancer diagnosis and prognosis;
KW tumorigenesis suppression.

XX Homo sapiens.

OS

XX US5352775-A.

XX PD 04-OCT-1994.

XX PF 08-AUG-1991; 91US-00741940.

XX PR 16-JAN-1991; 91GB-00000962.

XX PR 16-JAN-1991; 91GB-00000963.

XX PR 16-JAN-1991; 91GB-00000974.

XX PR 16-JAN-1991; 91GB-00000975.

XX XX

PA (ICIL) IMPERIAL CHEM IND.

PA (CANC-) CANCER INST.

PA (UYJO) UNIV JOHNS HOPKINS.

PA (UTAH) UNIV UTAH.

XX Markham AP, Hedge PJ, Anand R, Nakamura Y, Groden J, Kinzler K;

XX Thliveris A, Carlson M, Vogelstein B, Albertsen H, White RL;

XX Joslyn G;

XX WPI; 1994-316233/39.

XX DR N-PSDE; AAQ72297.

XX New human adenomatous polyposis coli DNA encoding tumour repressor - and

XX derived primers and probes for diagnosis, prognosis and treatment of

XX cancer.

XX Claim 3; Fig 7; 113pp; English.

XX AAQ72297 is a cDNA isolated from the human adenomatous polyposis coli

XX (APC) gene, it encodes the tumour repressors described in AAR63507 and

XX AAR63508. Determination of alterations in APC or its expression products,

XX can be used for the diagnosis and prognosis of cancer e.g. colorectal,

XX lung and breast tumours; and for determining predisposition to certain

CC cancers such as familial adenomatous polyposis (FAP) and Gardner's
CC syndrome. The wild type APC gene (or a part of it) can be used
CC therapeutically to restore gene function, while primers and probes
CC derived from the cDNA (AAQ72333-400 and AAQ72541-568) can be used to
CC detect mutations. Also APC proteins or analogues can be administered to
CC compensate for a defective gene, and epithelial cells, or transgenic
CC animals carrying a mutated APC allele are useful for detecting
CC therapeutic agents able to suppress tumorigenesis. (Updated on 25-MAR-
CC 2003 to correct PA field.)
XX
SQ Sequence 2860 AA;
Query Match 99.8%; Score 14551; DB 2; Length 2860;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2842; Conservative 0; Mismatches 0; Indels 18; Gaps 1;
QY 2 AAASYDOLLKQVEALKMNENLQCELEDNSNHLTKLETASNKMEVLKOLQGSIEDEANA 61
DB 1 AAASYDOLLKQVEALKMNENLQCELEDNSNHLTKLETASNKMEVLKOLQGSIEDEANA 60
QY 62 SSGQIDLLERLKEINLDS-----SNFFGVKLRKMSLRSGSREGSV 103
DB 61 SSGQIDLLERLKEINLDSNFFGVKLRKMSLRSGSREGSV 120
QY 104 SRSRGCSVPVPMGSPFRGFVNGSRESTGYLEBKERSILLADLKEEKEKDWYAAQLQ 163
DB 121 SRSRGCSVPVPMGSPFRGFVNGSRESTGYLEBKERSILLADLKEEKEKDWYAAQLQ 180
QY 164 NLTKRIIDSLPTENFSLQTLDTROLVEYARQIRVAMEEOLGTQCDMEKQARRIARIQ 223
DB 181 NLTKRIIDSLPTENFSLQTLDTROLVEYARQIRVAMEEOLGTQCDMEKQARRIARIQ 240
QY 224 IEKDILRIQLQSQATEARSSQKHETGSHDAERONEQGVGEINWATSGNGQSGTTR 283
DB 241 IEKDILRIQLQSQATEARSSQKHETGSHDAERONEQGVGEINWATSGNGQSGTTR 300
QY 284 MDHETASVSSSTHSPARLTHLGTKEVYVLLSMGLTHDKDMSRTLLAMSSQDS 343
DB 301 MDHETASVSSSTHSPARLTHLGTKEVYVLLSMGLTHDKDMSRTLLAMSSQDS 360
QY 344 CISMRQSCGLPLLQLLHGNDKDSVLLGNRSGSKSEARASAAALHNIHSPDDKRRRE 403
DB 361 CISMRQSCGLPLLQLLHGNDKDSVLLGNRSGSKSEARASAAALHNIHSPDDKRRRE 420
QY 404 IRVLHLLBOIRAYCETCEWQEAHEPGMDQKNPMPAPVEHQICPAVCVLKLSFDEHR 463
DB 421 IRVLHLLBOIRAYCETCEWQEAHEPGMDQKNPMPAPVEHQICPAVCVLKLSFDEHR 480
QY 464 HAMNELGLOAIABELLOVDCEMYGLTNDHYSITLRRVAGMALTNLTFGDVANKATLCMSK 523
DB 481 HAMNELGLOAIABELLOVDCEMYGLTNDHYSITLRRVAGMALTNLTFGDVANKATLCMSK 540
QY 524 GCMRALVAQLKSESEDLQOVVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKKES 583
DB 541 GCMRALVAQLKSESEDLQOVVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKKES 600
QY 584 TLKSVLSALWNLSAHCTENKADICAVDGLAFLVGLTYRSQTNLTALIESGGGILRNVS 643
DB 601 TLKSVLSALWNLSAHCTENKADICAVDGLAFLVGLTYRSQTNLTALIESGGGILRNVS 660
QY 644 SLIATNEDHROILRENNCLQTLLOLKHSHSLTIVSNACGTLWNLSARNPKDQOALWDMGA 703
DB 661 SLIATNEDHROILRENNCLQTLLOLKHSHSLTIVSNACGTLWNLSARNPKDQOALWDMGA 720
QY 704 VSMKXNLTHSKKMTAMGSAALRNLMANRPKAYKDANIMPGSLPSLHVKKOKALEAE 763
DB 721 VSMKXNLTHSKKMTAMGSAALRNLMANRPKAYKDANIMPGSLPSLHVKKOKALEAE 780
QY 764 LDAQHLSETFDNIDNLSPKASHRSKQHKQSLYGDYVFTNRHDDNRDNFTGMTVLS 823
DB 781 LDAQHLSETFDNIDNLSPKASHRSKQHKQSLYGDYVFTNRHDDNRDNFTGMTVLS 840
QY 824 PYLNTTVLPSSSSSRGSLDSSRSKDRSLERERIGLGNYPATENPGTSSKRGLOISTT 883

DB 841 PYLNTTVLPSSSSSRGSLDSSRSKDRSLERERIGLGNYPATENPGTSSKRGLOISTT 900
QY 884 AAQIAKMEEVSALHTSQEDSSSGSTTELHCVTDERNALRRSSAAHTSHNTYNTKSENS 943
DB 901 AAQIAKMEEVSALHTSQEDSSSGSTTELHCVTDERNALRRSSAAHTSHNTYNTKSENS 960
QY 944 NRTCSMPYAKLEYKRRSSNDLSNVSDNGYKRGOMKPSIESYEDDESKFCSCYQYPAD 1003
DB 961 NRTCSMPYAKLEYKRRSSNDLSNVSDNGYKRGOMKPSIESYEDDESKFCSCYQYPAD 1020
QY 1004 LAHKIHSANHMDDNDGELDTPIYNSLYKSYDFQLANSGRQSPQNERWARPKHIIIDEIKQS 1063
DB 1021 LAHKIHSANHMDDNDGELDTPIYNSLYKSYDFQLANSGRQSPQNERWARPKHIIIDEIKQS 1080
QY 1064 EQRQSRNQSTTVPVYTESTDDKHLKFQPHFQOQECVSPYRSGANGSETNRVSGNHGNIQ 1123
DB 1081 EQRQSRNQSTTVPVYTESTDDKHLKFQPHFQOQECVSPYRSGANGSETNRVSGNHGNIQ 1140
QY 1124 NVYSQLCEDDYEDDKPTNYSERYSEREEHHEERPTNYSTKYNEEKHVVQDPIDYSLKY 1183
DB 1141 NVYSQLCEDDYEDDKPTNYSERYSEREEHHEERPTNYSTKYNEEKHVVQDPIDYSLKY 1200
QY 1184 AYDIPISSQKQSFSSKSSGSSKTEHMSSESSENTTSSNAKRONQLHPSSAQSRSGOP 1243
DB 1201 AYDIPISSQKQSFSSKSSGSSKTEHMSSESSENTTSSNAKRONQLHPSSAQSRSGOP 1260
QY 1244 QAAATCKVSSINQETIOTYCYVEDTPICFRSCSSLSLSSAEDEICGNOTTOEADSANTLO 1303
DB 1261 QAAATCKVSSINQETIOTYCYVEDTPICFRSCSSLSLSSAEDEICGNOTTOEADSANTLO 1320
QY 1304 IABIKGKIGTRAEADPVSEVPAVSQHPRTKSRLOGSSLSSESARHKAVEFPFSGAKSPSK 1363
DB 1321 IABIKGKIGTRAEADPVSEVPAVSQHPRTKSRLOGSSLSSESARHKAVEFPFSGAKSPSK 1380
QY 1364 SGAQTPKSPPEHYVQETPLMFSCRTSVSSLSQSFESRSIASSVQSPSCGMVSGIISPSDL 1423
DB 1381 SGAQTPKSPPEHYVQETPLMFSCRTSVSSLSQSFESRSIASSVQSPSCGMVSGIISPSDL 1440
QY 1424 PDSPGOTMPPSSKTKPPPPQTAQTKREVKNKAPTAEKRESGPKQAAVNAAVQVQVLP 1483
DB 1441 PDSPGOTMPPSSKTKPPPPQTAQTKREVKNKAPTAEKRESGPKQAAVNAAVQVQVLP 1500
QY 1484 DADTLHAFATESTPDGFCSSLSALSALDEPFIQKDELRIIMPVQENDNGNETESEQPK 1543
DB 1501 DADTLHAFATESTPDGFCSSLSALSALDEPFIQKDELRIIMPVQENDNGNETESEQPK 1560
QY 1544 ESNENOKEAEKTIIDSEKOLLDDDDDDIEILEBCTIISAMPTKSRKGGKPAQTASKLPP 1603
DB 1561 ESNENOKEAEKTIIDSEKOLLDDDDDDIEILEBCTIISAMPTKSRKGGKPAQTASKLPP 1620
QY 1604 PVARKEPSQLPVYKLLPSONRLOPQKHVSFTPGDDMPRVYCVBGTPIINFSTATSLSLTIE 1663
DB 1621 PVARKEPSQLPVYKLLPSONRLOPQKHVSFTPGDDMPRVYCVBGTPIINFSTATSLSLTIE 1680
QY 1664 SPPELAAAGEVVRGGAQSGEFKRGDTIPTEGRSTDEAQQGKTSSVTIPELDDNKAEGDI 1723
DB 1681 SPPELAAAGEVVRGGAQSGEFKRGDTIPTEGRSTDEAQQGKTSSVTIPELDDNKAEGDI 1740
QY 1724 LAECINSAMPKGSHPKPRVKTIQMDVQQAASSASSAPNKNQDGGKKKPTSPVKPIPQNT 1783
DB 1741 LAECINSAMPKGSHPKPRVKTIQMDVQQAASSASSAPNKNQDGGKKKPTSPVKPIPQNT 1800
QY 1784 EYRTRVRKQADSKXNLNAERVFSDNPKDSKQNLKNNKDFNDKLPNNEDVRGSGFAFDP 1843
DB 1801 EYRTRVRKQADSKXNLNAERVFSDNPKDSKQNLKNNKDFNDKLPNNEDVRGSGFAFDP 1860
QY 1844 HHVTPTEGTPYCFSRNDSLSLDFDDDDVDLSREKAEELKAKENKESAKVTSHTLTSN 1903
DB 1861 HHVTPTEGTPYCFSRNDSLSLDFDDDDVDLSREKAEELKAKENKESAKVTSHTLTSN 1920
QY 1904 QQSANKTQAIKAPQINRGQPKPILOXQSTFPQSSKDIIPDRGAATDEKLOQFAJENTPVCF 1963

Db 1921 QOSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDIPDRGAATDBKLNFAIENTPVCF 1980
 Qy 1964 SHNSLSLSDIDQENNNKNEP:KETEPDPSQEPKPOASGYAPKSHFVEDTTPVCF 2023
 Db 1981 SHNSLSLSDIDQENNNKNEP:KETEPDPSQEPKPOASGYAPKSHFVEDTTPVCF 2040
 Qy 2024 NSLSLSLSDIDQECISSAMPKKPSRLKGNEXHSPNMGILGEDITLCLKDI 2083
 Db 2041 NSLSLSLSDIDQECISSAMPKKPSRLKGNEXHSPNMGILGEDITLCLKDI 2100
 Qy 2084 QRPDEHGLSPDENFDWKAIOGANSIVSSLHQAACLSQAASDSDSILSKSGIS 2143
 Db 2101 QRPDEHGLSPDENFDWKAIOGANSIVSSLHQAACLSQAASDSDSILSKSGIS 2160
 Qy 2144 LGPPFHLTPDOEKPFSTNKGPRILKPGEXSTLTETKIESESGIKGGKVKVKSITGV 2203
 Db 2161 LGPPFHLTPDOEKPFSTNKGPRILKPGEXSTLTETKIESESGIKGGKVKVKSITGV 2220
 Qy 2204 RSNSEISGOMKQPLQANMPSISRGRTMIHTPGVNSSSSTSPVSKGPPKLTSPAKSPSE 2263
 Db 2221 RSNSEISGOMKQPLQANMPSISRGRTMIHTPGVNSSSSTSPVSKGPPKLTSPAKSPSE 2280
 Qy 2264 GQTATTSRGAKPSVKSELSFVARQTSQIGGSSKAPSRSGSRDSTPSRPAQPLSRPIOS 2323
 Db 2281 GQTATTSRGAKPSVKSELSFVARQTSQIGGSSKAPSRSGSRDSTPSRPAQPLSRPIOS 2340
 Qy 2324 PGRNSIPGRNGISPPNKLSQLPSTSPSTASTKSSGSGKMSYTPSGROMSQOQLTKQTG 2383
 Db 2341 PGRNSIPGRNGISPPNKLSQLPSTSPSTASTKSSGSGKMSYTPSGROMSQOQLTKQTG 2400
 Qy 2384 LSKNASSIPRESASKGLNOMNGANKVLSRMSSTKSSGSESDRSERPVLRQSTF 2443
 Db 2401 LSKNASSIPRESASKGLNOMNGANKVLSRMSSTKSSGSESDRSERPVLRQSTF 2460
 Qy 2444 IKZAPPTLRKLEESASFESISPSRSPASPTRSQATPVLSPSLPDMSLSTHSSVQAGG 2503
 Db 2461 IKZAPPTLRKLEESASFESISPSRSPASPTRSQATPVLSPSLPDMSLSTHSSVQAGG 2520
 Qy 2504 WRKLPNLSTIYDGRPAKRDIAHSHSESPSRPLPINRSGTWKREKSHSSSLPRVST 2563
 Db 2521 WRKLPNLSTIYDGRPAKRDIAHSHSESPSRPLPINRSGTWKREKSHSSSLPRVST 2580
 Qy 2564 WRTGSSSILSASSSESEKAKSEDEKHVNSISGKQSKENOVSAKGTWRKIKENEFSP 2623
 Db 2581 WRTGSSSILSASSSESEKAKSEDEKHVNSISGKQSKENOVSAKGTWRKIKENEFSP 2640
 Qy 2624 NSTSQTVSSGATNGABESKTLIYQMAFAVSKTEDVWVRIEDCPINNPRSGRSPGTGTPVI 2683
 Db 2641 NSTSQTVSSGATNGABESKTLIYQMAFAVSKTEDVWVRIEDCPINNPRSGRSPGTGTPVI 2700
 Qy 2684 DSVSEKANPNIKDKONQAKQNVGSGVPMRTVGLNRLTFSIQVADPDQKTEIKPGON 2743
 Db 2701 DSVSEKANPNIKDKONQAKQNVGSGVPMRTVGLNRLTFSIQVADPDQKTEIKPGON 2760
 Qy 2744 NPVPVSETHESIVERTPFSSSSSKHSPSPGTVAARTVPPNPNPSPKSGADSTSAAPS 2803
 Db 2761 NPVPVSETHESIVERTPFSSSSSKHSPSPGTVAARTVPPNPNPSPKSGADSTSAAPS 2820
 Qy 2804 QIPTPVNNNTKRDSDKTDSTESSGTSQPKRHSGSVLVTSV 2843
 Db 2821 QIPTPVNNNTKRDSDKTDSTESSGTSQPKRHSGSVLVTSV 2860

RESULT 6
 AAW76140
 ID AAW76140 standard; protein; 2843 AA.
 XX
 AC AAW76140;
 XX
 DT 25-MAR-2003 (revised)
 DT 23-NOV-1998 (first entry)
 XX
 DE Human APC protein #1.

XX Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
 KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
 KW Gardner's Syndrome; GS; predisposition.
 XX
 OS Homo sapiens.
 XX US5783666-A.
 XX 21-JUL-1998.
 XX 25-MAY-1995; 95US-00452655.
 XX 16-JAN-1991; 91GB-00000962.
 XX 16-JAN-1991; 91GB-00000963.
 XX 16-JAN-1991; 91GB-00000974.
 XX 16-JAN-1991; 91GB-00000975.
 XX 08-AUG-1991; 91US-00741940.
 XX 12-AUG-1994; 94US-00289548.
 XX (CANC-) CANCER INST.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UTAH) UNIV UTAH.
 PA (ZENE) ZENECA PHARM.
 XX Kinzler K, Joslyn G, Markham AP, Carlson M, White RL;
 PI Thiliveris A, Groden J, Anand R, Nakamura Y, Albertsen H;
 PI Vogelstein B, Hedge PJ;
 XX WPI: 1998-427100/36.
 DR N-PSDB; AAV56447.
 XX
 PT Adenomatous polyposis coli protein - useful in the treatment of cancers
 associated with mutation(s) on human chromosome 5q21.
 XX
 PS Disclosure; Col 41-54; 102pp; English.
 XX
 CC This sequence represents a human familial adenomatous polyposis coli
 (APC) protein from clone DP2.5. The gene for the protein is present on
 human chromosome 5q21 and is also referred to as adenomatous polyposis
 coli gene. It is a tumour suppressor gene, and mutations in this gene
 have been associated with tumorigenesis in retinoblastoma and colorectal
 tumours, and especially with familial adenomatous polyposis (FAP) and
 Gardner's Syndrome (GS). The protein can be used in therapy to replace
 lack of native functional protein and the nucleic acids can be used for
 gene therapy. The nucleic acids that encode them can also be used as
 probes and primers in detection of the cancers and predisposition to it.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC
 SQ Sequence 2843 AA;

Query Match 99.7%; Score 14533; DB 2; Length 2843;
 Best Local Similarity 99.8%; Pred. NO. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 MAAASYDQLLKQVEALKMENSLRQLEDNSNHLTKLETEASNKMKVLLKQLOQSIIDEAM 60
 Db 1 MAAASYDQLLKQVEALKMENSLRQLEDNSNHLTKLETEASNKMKVLLKQLOQSIIDEAM 60
 Qy 61 ASSGQIDLLERLKEINLDSNPFVGLRSMKSLRSYSGREGSVSSRGCSVPVWGSFPR 120
 Db 61 ASSGQIDLLERLKEINLDSNPFVGLRSMKSLRSYSGREGSVSSRGCSVPVWGSFPR 120
 Qy 121 RGFVNGSRSTGYLEBELEKERSILLADLDKEKEKDWYVAQLQNLTKRIDSPLTENFSL 180
 Db 121 RGFVNGSRSTGYLEBELEKERSILLADLDKEKEKDWYVAQLQNLTKRIDSPLTENFSL 180
 Qy 181 QDTLTRQLEYEARQIRVAMEBQLGTCQDMEKRAQRIARIQOIKDILIRQLIQSQAT 240
 Db 181 QDTLTRQLEYEARQIRVAMEBQLGTCQDMEKRAQRIARIQOIKDILIRQLIQSQAT 240
 Qy 241 EAERSQNKHETGSHDAERQNEGQGVGEINMATSNGOGSTTRMDHETASVLSSTSTSA 300

Db 241 EAERSQNKHETGSHDAERQNEGQVGEINMATSCNGOGSTTRMDHETASVLSSESSTSA 300
Qy 301 PRLTSLHGTKYEMVYVILLSMLGTHDKDDMDERTLLAMSSODSCISWRQSGCLPLLQLL 360
Db 301 PRLTSLHGTKYEMVYVILLSMLGTHDKDDMDERTLLAMSSQDSCISWRQSGCLPLLQLL 360
Qy 361 HGNDKDSVLLGNSRGSKARARASAALENIIHSQDDDKRGREIRVLLHLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNSRGSKARARASAALENIIHSQDDDKRGREIRVLLHLEQIRAYCETC 420
Qy 421 WQQAHAPEGMDQKNPMPAPVEHOICPAVCVLMKLSFDEBHRHAMNELGGLQIAIELLO 480
Db 421 WQQAHAPEGMDQKNPMPAPVEHOICPAVCVLMKLSFDEBHRHAMNELGGLQIAIELLO 480
Qy 481 VDCMYGLTNDHYSITLRYAGMALTNLTFFGVANKATLCSMKGCMRALVAQLKSESEDL 540
Db 481 VDCMYGLTNDHYSITLRYAGMALTNLTFFGVANKATLCSMKGCMRALVAQLKSESEDL 540
Qy 541 QOVIASVLNLSWRADVNSKTLTRVGSVKALMECALEVKKESTLKVLSALMNLSAHCT 600
Db 541 QOVIASVLNLSWRADVNSKTLTRVGSVKALMECALEVKKESTLKVLSALMNLSAHCT 600
Qy 601 ENKADICAVDGAFLVGLTVYRSQNTNLTATIESGGGILRVNVSLLIATNEDHROILRENN 660
Db 601 ENKADICAVDGAFLVGLTVYRSQNTNLTATIESGGGILRVNVSLLIATNEDHROILRENN 660
Qy 661 CLQTLLOHLKSHSLTVSNACOTLNLGARNPKOQALWDMGVMSMLKNLHSHKHMIAM 720
Db 661 CLQTLLOHLKSHSLTVSNACOTLNLGARNPKOQALWDMGVMSMLKNLHSHKHMIAM 720
Qy 721 GSAALRNLMANRPKAKYKXANIMSPGSSLSPLSHVRKOKALEDAQHLSETPDNIOLNS 780
Db 721 GSAALRNLMANRPKAKYKXANIMSPGSSLSPLSHVRKOKALEDAQHLSETPDNIOLNS 780
Qy 781 PKASHRSKORHKQSLYGDYVFTNRRHNDNRDNFTGNMTVLSPLNTTVLPSSSSSRGS 840
Db 781 PKASHRSKORHKQSLYGDYVFTNRRHNDNRDNFTGNMTVLSPLNTTVLPSSSSSRGS 840
Qy 841 LBSRSSEKDRSLERBERGICLGNHYHPATENPGTSSKRGLOISTTAAQIAKWEEVSAJHTS 900
Db 841 LBSRSSEKDRSLERBERGICLGNHYHPATENPGTSSKRGLOISTTAAQIAKWEEVSAJHTS 900
Qy 901 QEDRSSGSTTEHLVCTVDERNALRRSSAAHTSNYNTFTKSENSNRCTCMPYAKLEYKRSS 960
Db 901 QEDRSSGSTTEHLVCTVDERNALRRSSAAHTSNYNTFTKSENSNRCTCMPYAKLEYKRSS 960
Qy 961 NDLANSVSSNDGYKRGQWKPSIESYSEDDSKFCYQYPADLAHAKHSHANHMDNDGE 1020
Db 961 NDLANSVSSNDGYKRGQWKPSIESYSEDDSKFCYQYPADLAHAKHSHANHMDNDGE 1020
Qy 1021 LPTPINYSKYDEQLNSGRQSPSQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Db 1021 LPTPINYSKYDEQLNSGRQSPSQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFPQHFQOECVPSYRSGANGSETNRVGSNHGINQVSOQLCOEDDYEDDKP 1140
Db 1081 STDDKHLKFPQHFQOECVPSYRSGANGSETNRVGSNHGINQVSOQLCOEDDYEDDKP 1140
Qy 1141 TNYSERYSSEEQHEEBEERTNYSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSFSEKS 1200
Db 1141 TNYSERYSSEEQHEEBEERTNYSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSFSEKS 1200
Qy 1201 SSCQSKTTHMSSSENTSTPSSNAKRONOLHPSAOSRSCOPAKATCKVSSINQSTIQ 1260
Db 1201 SSCQSKTTHMSSSENTSTPSSNAKRONOLHPSAOSRSCOPAKATCKVSSINQSTIQ 1260
Qy 1261 TYCVEDTPICFRCSLSLSAEDEIGCNQTOEADSAANTLQIAEIKGKITGTRSAEDPV 1320
Db 1261 TYCVEDTPICFRCSLSLSAEDEIGCNQTOEADSAANTLQIAEIKGKITGTRSAEDPV 1320
Qy 1321 SEVPASVOHPRTKSRLOCGSSLSSESARHKAVEFPSSGAKSPKSGAQTPKSPPEHYVOET 1380
Db 1321 SEVPASVOHPRTKSRLOCGSSLSSESARHKAVEFPSSGAKSPKSGAQTPKSPPEHYVOET 1380

Qy 1381 PLMFRCCTSVSLSLDSFBSRSIASSVQSEPCSGMVSGIISPSDLDPSPCQOTMPPSRSKTTP 1440
Db 1381 PLMFRCCTSVSLSLDSFBSRSIASSVQSEPCSGMVSGIISPSDLDPSPCQOTMPPSRSKTTP 1440
Qy 1441 PPQTAQTKREVPKNKAPTAEKRESGPKQAANAQVQVLPADTLLHPATESTPDGF 1500
Db 1441 PPQTAQTKREVPKNKAPTAEKRESGPKQAANAQVQVLPADTLLHPATESTPDGF 1500
Qy 1501 SCSSLSLSALSDPEPIQKDVLEIRIMPPVQENDNGNETESBOPKESNENQEKAKTIDSE 1560
Db 1501 SCSSLSLSALSDPEPIQKDVLEIRIMPPVQENDNGNETESBOPKESNENQEKAKTIDSE 1560
Qy 1561 KDLDDSDDDDIIELEECIIISAMPTKSRKKGKPAQATSKLPPPVARKPSPOLPVYKLLPS 1620
Db 1561 KDLDDSDDDDIIELEECIIISAMPTKSRKKGKPAQATSKLPPPVARKPSPOLPVYKLLPS 1620
Qy 1621 QNRLOPQKHVSFTPGDDMPRYVCVEGTPINFSTATSLDLTIESPPNELAAGEVGRGAQ 1680
Db 1621 QNRLOPQKHVSFTPGDDMPRYVCVEGTPINFSTATSLDLTIESPPNELAAGEVGRGAQ 1680
Qy 1681 SGEFEKRDITPTEGRSTDEAOGGKTSSVTIPELDDNKAEEGDILAEICINSAMPKKGKSHKP 1740
Db 1681 SGEFEKRDITPTEGRSTDEAOGGKTSSVTIPELDDNKAEEGDILAEICINSAMPKKGKSHKP 1740
Qy 1741 FRVKIMQVOQASASSAPNKNQLDGKKKPTSPVPIPONTYRTRVRKNADSKNMLN 1800
Db 1741 FRVKIMQVOQASASSAPNKNQLDGKKKPTSPVPIPONTYRTRVRKNADSKNMLN 1800
Qy 1801 ARVFSDNKKSKQNLKNNSKDFNDKLPNNEDRVGRGFAPDSPHHYTPIEGTPICFSRND 1860
Db 1801 ARVFSDNKKSKQNLKNNSKDFNDKLPNNEDRVGRGFAPDSPHHYTPIEGTPICFSRND 1860
Qy 1861 SLSSLDFFDDDDVLDLSREKAEILKAKENKESAKVTSTELTSSNOOSANKTOAIKQPINR 1920
Db 1861 SLSSLDFFDDDDVLDLSREKAEILKAKENKESAKVTSTELTSSNOOSANKTOAIKQPINR 1920
Qy 1921 GQPKILQKQSTFPPOSSKDIIDRGAATDEKLQNEFAIENTPVCFSHNSLSLSLSDIDENN 1980
Db 1921 GQPKILQKQSTFPPOSSKDIIDRGAATDEKLQNEFAIENTPVCFSHNSLSLSLSDIDENN 1980
Qy 1981 NKNEPFIKETBPPDSQGEPSKPQASGVAPKHFVEDTPVCFSRNSSISSLSIDSEDLQ 2040
Db 1981 NKNEPFIKETBPPDSQGEPSKPQASGVAPKHFVEDTPVCFSRNSSISSLSIDSEDLQ 2040
Qy 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
Db 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
Qy 2101 WPAIPEGANSIVSSLUHQAAAACISROASSSDSILSKGSIILGSPFHLTPDQEEKPFT 2160
Db 2101 WPAIPEGANSIVSSLUHQAAAACISROASSSDSILSKGSIILGSPFHLTPDQEEKPFT 2160
Qy 2161 SNKGPRILKPEKSTLETKTESKGIKGGKVVYKSLITGKVRNSGEISQOMKQLOAN 2220
Db 2161 SNKGPRILKPEKSTLETKTESKGIKGGKVVYKSLITGKVRNSGEISQOMKQLOAN 2220
Qy 2221 MPISIRGRMTIHIPVRNNSSTSPVSKGPPKLTTPASKSPSEGTATTSPRGAKPVKS 2280
Db 2221 MPISIRGRMTIHIPVRNNSSTSPVSKGPPKLTTPASKSPSEGTATTSPRGAKPVKS 2280
Qy 2281 ELSPVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPTQSPGRNSISPRNGISPPN 2340
Db 2281 ELSPVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPTQSPGRNSISPRNGISPPN 2340
Qy 2341 KLSQLPRTSSPTASTKSGSGKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSSPTASTKSGSGKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASKG 2400
Qy 2401 LNMKNNGANKKVELSRMSTKSGSSEDRSPVLVROSTFIKEAPSPFLRKLBSA 2460
Db 2401 LNMKNNGANKKVELSRMSTKSGSSEDRSPVLVROSTFIKEAPSPFLRKLBSA 2460

QY 2461 SFESLSPSSRPASPTRSOATPVLSPLPMSLSTHSSVOAGGWRKLPNLSPTIYNDG 2520
DB 2461 SFESLSPSSRPASPTRSOATPVLSPLPMSLSTHSSVOAGGWRKLPNLSPTIYNDG 2520
QY 2521 RPAKRHHIARSHSPPRLPINSRGTWKRSHKSHSSLPVSTWRTGSSSSILSASSES 2580
DB 2521 RPAKRHHIARSHSPPRLPINSRGTWKRSHKSHSSLPVSTWRTGSSSSILSASSES 2580
QY 2581 SEKAKSDEKHVNSISGTSKQENQVSAKGTWKIKENEFSPNSTQTVSSGATNGAES 2640
DB 2581 SEKAKSDEKHVNSISGTSKQENQVSAKGTWKIKENEFSPNSTQTVSSGATNGAES 2640
QY 2641 KTLIIQWAPAVSKTEDVWVIEDCPINNPGRSPGTGNTPPVIDSVSEKAPNPKDSKDN 2700
DB 2641 KTLIIQWAPAVSKTEDVWVIEDCPINNPGRSPGTGNTPPVIDSVSEKAPNPKDSKDN 2700
QY 2701 QAKQNVGNGSVPMRTVGLNRLTSFTQVDAPDQKGTETKPGQNNPVPVSETNESPIVERT 2760
DB 2701 QAKQNVGNGSVPMRTVGLNRLTSFTQVDAPDQKGTETKPGQNNPVPVSETNESPIVERT 2760
QY 2761 PFSSSSSSKHSFSGTVAARVTFPNYNPSPRKSADSTSGARPSQIPTPVNNNTKRDST 2820
DB 2761 PFSSSSSSKHSFSGTVAARVTFPNYNPSPRKSADSTSGARPSQIPTPVNNNTKRDST 2820
QY 2821 DSTEESGTQSPKHSGLYTSV 2843
DB 2821 DSTEESGTQSPKHSGLYTSV 2843

RESULT 7
AAW76144
ID AAW76144 standard; protein; 2843 AA.
AC AAW76144;
XX
XX
DT 25-MAR-2003 (revised)
DT 23-NOV-1998 (first entry)
DE Human APC protein #2.
XX
XX Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
KW Gardner's syndrome; GS; predisposition.
XX
OS Homo sapiens.
XX
XX US5783666-A.
XX
XX 21-JUL-1998.
XX
XX 25-MAY-1995; 95US-00452655.
XX
XX 16-JAN-1991; 91GB-00000962.
XX 16-JAN-1991; 91GB-00000963.
XX 16-JAN-1991; 91GB-00000974.
XX 16-JAN-1991; 91GB-00000975.
XX 08-AUG-1991; 91US-00741940.
XX 12-AUG-1994; 94US-00289548.
XX
XX (CANC-) CANCER INST.
XX (UYJO) UNIV JOHNS HOPKINS.
XX (UTAH) UNIV UTAH.
XX (ZENE) ZENECA PHARM.
XX
XX Kinzler X, Joslyn G, Markham AF, Carlson M, White RL;
PI Thliveris A, Groden J, Nakamura Y, Albertsen H;
PI Vogelstein B, Hedge PJ;
PI
XX WPI; 1998-427100/36.
XX
XX Adenomatous polyposis coli protein - useful in the treatment of cancers
XX associated with mutation(s) on human chromosome 5q21
XX

PS Disclosure; Col 63-78; 102pp; English.
XX
CC This sequence represents a human familial adenomatous polyposis coli
CC (APC) protein isolated from 87 cDNA clones. The gene for the protein is
CC present on human chromosome 5q21 and is also referred to as adenomatous
CC polyposis coli gene. It is a tumour suppressor gene, and mutations in
CC this gene have been associated with tumorigenesis in retinoblastoma and
CC colorectal tumours, and especially familial adenomatous polyposis (FAP)
CC and Gardner's syndrome (GS). The protein can be used in therapy to
CC replace lack of native functional protein and the nucleic acids can be
CC used for gene therapy. The nucleic acids that encode them can also be
CC used as probes and primers in detection of the cancers and predisposition
CC to it. (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2843 AA;

Query Match 99.7%; Score 14533; DB 2; Length 2843;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAASYDOLLKQVEALQWENSLRQLEEDNSNHLTKLETEASNKKEVYLKQLOQSIDEAM 60
DB 1 MAAASYDOLLKQVEALQWENSLRQLEEDNSNHLTKLETEASNKKEVYLKQLOQSIDEAM 60
QY 61 ASSQOIDLRLKELNLDSSNFPVGLRSMVLSRYSYSGREGSVSSSGECSVPVPMGSFPR 120
DB 61 ASSQOIDLRLKELNLDSSNFPVGLRSMVLSRYSYSGREGSVSSSGECSVPVPMGSFPR 120
QY 121 RGFVNGSRESTGYLEELKERSLLLDLDKEEKEKDWYLAQLQNLTKRIDSPLTENFSL 180
DB 121 RGFVNGSRESTGYLEELKERSLLLDLDKEEKEKDWYLAQLQNLTKRIDSPLTENFSL 180
QY 181 QTLTTRQLEVEARQIRVAMEEOLGTCQDMKEKRAQRIARIQOIEKDIIRIROLLOSQT 240
DB 181 QTLTTRQLEVEARQIRVAMEEOLGTCQDMKEKRAQRIARIQOIEKDIIRIROLLOSQT 240
QY 241 EBERSQNHETGSHDAERQNEQGVGEINMATSNGQGQSTTRMDHETASVLSSSSTHSA 300
DB 241 EBERSQNHETGSHDAERQNEQGVGEINMATSNGQGQSTTRMDHETASVLSSSSTHSA 300
QY 301 PRLTSLHGTGVEMVYSLISMLGTHDKDMSRTLLAMSSSQSCISMROSQGLPLLIQLL 360
DB 301 PRLTSLHGTGVEMVYSLISMLGTHDKDMSRTLLAMSSSQSCISMROSQGLPLLIQLL 360
QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHISQPDQKGRREIRVHLLEQIRAYCETC 420
DB 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHISQPDQKGRREIRVHLLEQIRAYCETC 420
QY 421 WEMQEAHEPGMDQKNPMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGGLOAIAELIQ 480
DB 421 WEMQEAHEPGMDQKNPMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGGLOAIAELIQ 480
QY 481 VDCMYGLTNDHYSITLRRYAGMALTNLTFFGVANKATLCSMKGCWRALVAOLKSESED 540
DB 481 VDCMYGLTNDHYSITLRRYAGMALTNLTFFGVANKATLCSMKGCWRALVAOLKSESED 540
QY 541 QQVIASVRLNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVALSWLNLSAHT 600
DB 541 QQVIASVRLNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVALSWLNLSAHT 600
QY 601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILRNVSLLIATNEDHRQLRENN 660
DB 601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILRNVSLLIATNEDHRQLRENN 660
QY 661 CLQTLQLKSHSLTIVSNACOTLWNLARNPKDQDALWDMGAVSMLKXLIHSHKMTAM 720
DB 661 CLQTLQLKSHSLTIVSNACOTLWNLARNPKDQDALWDMGAVSMLKXLIHSHKMTAM 720
QY 721 GSAALRNLMANRPKYKDANIMSPGSSLPVHVKQKALEAELDAQHLSHSETFNIDLS 780
DB 721 GSAALRNLMANRPKYKDANIMSPGSSLPVHVKQKALEAELDAQHLSHSETFNIDLS 780

Db 781 PKASHRSKORHKSGLYGDYFDTVRHDDNRSDNFTGNMTVLSPLYLNTTVLPSSSSRGSS 840
Qy 841 LDSRSSEKDRSLERERGI GLGNYPATENPGTSSKRGQLQISTTAAQIAKVMEEVSALHTS 900
Db 841 LDSRSSEKDRSLERERGI GLGNYPATENPGTSSKRGQLQISTTAAQIAKVMEEVSALHTS 900
Qy 901 QEDRSSGGTTLHLCTVDERNALRSSAAHTSNTYNTFKSENSNRTCSMPYAKULEYKRSS 960
Db 901 QEDRSSGGTTLHLCTVDERNALRSSAAHTSNTYNTFKSENSNRTCSMPYAKULEYKRSS 960
Qy 961 NDSLNSVSSNDGYKRGOMKPSIESYSBDDSKFCVSGQYPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLNSVSSNDGYKRGOMKPSIESYSBDDSKFCVSGQYPADLAHKIHSANHMDNDGE 1020
Qy 1021 LDTPIYNSLYKSDOLNSGROSPONERWARPKHI IEDEIKQSEOROSNQSTTYPVYTE 1080
Db 1021 LDTPIYNSLYKSDOLNSGROSPONERWARPKHI IEDEIKQSEOROSNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFPQHFQGOECVSPYRERGANGSTNRVGNHGINQVNSQSLCOEDDYEDDKP 1140
Db 1081 STDDKHLKFPQHFQGOECVSPYRERGANGSTNRVGNHGINQVNSQSLCOEDDYEDDKP 1140
Qy 1141 TNYSERYSSEEEHQBEEERTNYISIKYNEEKHVPDQPIDYSLKYATDIPSSQKQSFSPKS 1200
Db 1141 TNYSERYSSEEEHQBEEERTNYISIKYNEEKHVPDQPIDYSLKYATDIPSSQKQSFSPKS 1200
Qy 1201 SSGOSSKTEHSSSENSTPSSNAKQNLHPSSAQSRSGOPOKAATCKVSSINQETIQ 1260
Db 1201 SSGOSSKTEHSSSENSTPSSNAKQNLHPSSAQSRSGOPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTPICFSRCSLSLSAASDEIGCNQTTQEADSANTLQIAEIKKIGTISAEDPV 1320
Db 1261 TYCVEDTPICFSRCSLSLSAASDEIGCNQTTQEADSANTLQIAEIKKIGTISAEDPV 1320
Qy 1321 SEVPAVSQHPRTKSLRLOGSLSESARHKAVERPSSGAKSPSKSCAQTPKSPPHYVOET 1380
Db 1321 SEVPAVSQHPRTKSLRLOGSLSESARHKAVERPSSGAKSPSKSCAQTPKSPPHYVOET 1380
Qy 1381 PLMFSRCTSVSLDSFEBSRTASSVQSEPCSGMWGIIISPSDLDPSPGQTPPERSKTTP 1440
Db 1381 PLMFSRCTSVSLDSFEBSRTASSVQSEPCSGMWGIIISPSDLDPSPGQTPPERSKTTP 1440
Qy 1441 PPPOTAQTKREVPKNKAPTAKRBSGPKQAANAAVQVQVLPDADTLHPFATESTPDGF 1500
Db 1441 PPPOTAQTKREVPKNKAPTAKRBSGPKQAANAAVQVQVLPDADTLHPFATESTPDGF 1500
Qy 1501 SCSSLSALSLEDEPIQKDVLRIMPPVOENDNGNETESEOPKESNENQKEAEKTIIDSE 1560
Db 1501 SCSSLSALSLEDEPIQKDVLRIMPPVOENDNGNETESEOPKESNENQKEAEKTIIDSE 1560
Qy 1561 KDLLDDSDDDDIIELEECIIISAMPTKSSRKKGKPAQATASKLPPPVARKPSOLPYVKLLPS 1620
Db 1561 KDLLDDSDDDDIIELEECIIISAMPTKSSRKKGKPAQATASKLPPPVARKPSOLPYVKLLPS 1620
Qy 1621 QNRLOPQKHVSFTPGDDDMPRVYCVEGTPIINFSTATISLSDLTIESPPNELAAGEVGRGAQ 1680
Db 1621 QNRLOPQKHVSFTPGDDDMPRVYCVEGTPIINFSTATISLSDLTIESPPNELAAGEVGRGAQ 1680
Qy 1681 SGEFEKRTIPIEGRSTDEAOGKTSSTVTIPELDDNKAEEGDIIAEACINSAMPKGGHKP 1740
Db 1681 SGEFEKRTIPIEGRSTDEAOGKTSSTVTIPELDDNKAEEGDIIAEACINSAMPKGGHKP 1740
Qy 1741 FRVKKIMDQVOQASASSAPKNQLDGKKKXPTSPVKPIPQNTYRTRVRKNADSKNLN 1800
Db 1741 FRVKKIMDQVOQASASSAPKNQLDGKKKXPTSPVKPIPQNTYRTRVRKNADSKNLN 1800
Qy 1801 AERVFSDNKDSKKQNLKNNKDFNDKLPNNEDRVGSPAFDSPHHYTPIEGTPYCFSRND 1860
Db 1801 AERVFSDNKDSKKQNLKNNKDFNDKLPNNEDRVGSPAFDSPHHYTPIEGTPYCFSRND 1860
Qy 1861 SLSSLDFFDDDDVLSREKAELRKAKENKESAKVTSHTELTSNQOSANKTOIAIAKQPINR 1920

Db 1861 SLSSLDFFDDDDVLSREKAELRKAKENKESAKVTSHTELTSNQOSANKTOIAIAKQPINR 1920
Qy 1921 GQPKPILOKQSTFPOSSKDIPIPRGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Db 1921 GQPKPILOKQSTFPOSSKDIPIPRGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Qy 1981 KNEBEDIKETEPPDDSGQEPSKQASGYAPKSPHVEDTPVCFSRNSSLSLSLSDIDQENN 2040
Db 1981 KNEBEDIKETEPPDDSGQEPSKQASGYAPKSPHVEDTPVCFSRNSSLSLSLSDIDQENN 2040
Qy 2041 ECITSSAMPKKKPSRLKGDNEKHSFRNMGCIIGEDITLDLKDIQRPDSEHLSPOSENED 2100
Db 2041 ECITSSAMPKKKPSRLKGDNEKHSFRNMGCIIGEDITLDLKDIQRPDSEHLSPOSENED 2100
Qy 2101 WKAIQEGANSIVSSLHQAAAAACLRSQASSDSDSLSLKSGISLGSPPFHLTPDQBEKPT 2160
Db 2101 WKAIQEGANSIVSSLHQAAAAACLRSQASSDSDSLSLKSGISLGSPPFHLTPDQBEKPT 2160
Qy 2161 SNKGPRILAPGKSTLETTKIESESGIKGGKVKYKSLITGKVRNSSEISGOMKQPLQAN 2220
Db 2161 SNKGPRILAPGKSTLETTKIESESGIKGGKVKYKSLITGKVRNSSEISGOMKQPLQAN 2220
Qy 2221 MFSISGRGTMIHIPGVNRSSSTSPVSKGPPKLPKTPASKPSEGGQTATTSPRGAKEPVSKS 2280
Db 2221 MFSISGRGTMIHIPGVNRSSSTSPVSKGPPKLPKTPASKPSEGGQTATTSPRGAKEPVSKS 2280
Qy 2281 ELSPVARQTSQIGGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPGRNGISPPN 2340
Db 2281 ELSPVARQTSQIGGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSPSTASTKSSGSGKMYTSPGRQMSQOQLTKOTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSPSTASTKSSGSGKMYTSPGRQMSQOQLTKOTGLSKNASSIPRSESASKG 2400
Qy 2401 LNQMNGNCAKVKYELSRMSSYKSSGSESDRSERPVLVQSTFIKEAPSTLRRKLEESA 2460
Db 2401 LNQMNGNCAKVKYELSRMSSYKSSGSESDRSERPVLVQSTFIKEAPSTLRRKLEESA 2460
Qy 2461 SFESLSPPSRPASPTRSQAQTPLVPSLSDMSLSTHSSVQAGGWRKLPENLSPTIEYNDG 2520
Db 2461 SFESLSPPSRPASPTRSQAQTPLVPSLSDMSLSTHSSVQAGGWRKLPENLSPTIEYNDG 2520
Qy 2521 RPAKRHDIAARSHSESPRLPINRSGTWKREHKHSSSLPRVSTWRTTSGSSSILSASSES 2580
Db 2521 RPAKRHDIAARSHSESPRLPINRSGTWKREHKHSSSLPRVSTWRTTSGSSSILSASSES 2580
Qy 2581 SEKAKSEDEKHVNSISGTTKOSKENOVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKHVNSISGTTKOSKENOVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
Qy 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKDN 2700
Db 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKDN 2700
Qy 2701 QAKQNYGNGSVPMRTVGLNRLTSFTQVDAPOKGTETIKPGQNNPVPVSETNESPVERT 2760
Db 2701 QAKQNYGNGSVPMRTVGLNRLTSFTQVDAPOKGTETIKPGQNNPVPVSETNESPVERT 2760
Qy 2761 PFSSSSSSKHSFSGTVAARVTPFNTNPSFRKSSADTSARPSQIPTPVNNNTKKEDSKT 2820
Db 2761 PFSSSSSSKHSFSGTVAARVTPFNTNPSFRKSSADTSARPSQIPTPVNNNTKKEDSKT 2820
Qy 2821 DSTESSGTQSPKHSGLYVTSV 2843
Db 2821 DSTESSGTQSPKHSGLYVTSV 2843

RESULT 8
AAB23011
ID AAB23011 standard; protein; 2843 AA.
XX
XX AAB23011;
XX

Db 1021 LDTPIVSLKYSDBQLNSGRQSPQNERWAPKHHIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFOHFQOQCVSPYRGRGANGSETNRVGNHGINQVNSOSLQCEDDYEDDKP 1140
Db 1081 STDDKHLKFOHFQOQCVSPYRGRGANGSETNRVGNHGINQVNSOSLQCEDDYEDDKP 1140
Qy 1141 TNYSERVSEEEHQBEEERPTNYSIKYNEBKHXVDQPIDYSLKYATDIPSSQKQSFSEFSK 1200
Db 1141 TNYSERVSEEEHQBEEERPTNYSIKYNEBKHXVDQPIDYSLKYATDIPSSQKQSFSEFSK 1200
Qy 1201 SSGQSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQETIQ 1260
Db 1201 SSGQSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTPICFSRCSLSSLSAASDETCNQTTQEAADSANTLQIAETKIGKIGTRSAEDPV 1320
Db 1261 TYCVEDTPICFSRCSLSSLSAASDETCNQTTQEAADSANTLQIAETKIGKIGTRSAEDPV 1320
Qy 1321 SEVPAVSHQPTKSRRLQSSLSSESARHKAVERPSSGAKSPSKCAOTPKSPPEHYVOET 1380
Db 1321 SEVPAVSHQPTKSRRLQSSLSSESARHKAVERPSSGAKSPSKCAOTPKSPPEHYVOET 1380
Qy 1381 PLMFSRCTSVSSLSDFSRSTASSVQSEPCSGMVGSIISPSDLDPSPGQTMPPSRSKTTP 1440
Db 1381 PLMFSRCTSVSSLSDFSRSTASSVQSEPCSGMVGSIISPSDLDPSPGQTMPPSRSKTTP 1440
Qy 1441 PPQTAOTKREVPKNKAPTAEKRESGPKQAANVAQVQVLPDADTLLHFAESTPDGF 1500
Db 1441 PPQTAOTKREVPKNKAPTAEKRESGPKQAANVAQVQVLPDADTLLHFAESTPDGF 1500
Qy 1501 SCSSLSALSALDEFFIOKQVELRIMPVQENDNGNETESEQPKESNENQKEAEKTTIDSE 1560
Db 1501 SCSSLSALSALDEFFIOKQVELRIMPVQENDNGNETESEQPKESNENQKEAEKTTIDSE 1560
Qy 1561 KOLLDDSDDDIEILEECIISAMTKSRKKGKPAQATASKLPPVARKPQOLPVYKLLPS 1620
Db 1561 KOLLDDSDDDIEILEECIISAMTKSRKKGKPAQATASKLPPVARKPQOLPVYKLLPS 1620
Qy 1621 QNRLOPKQHVSTFGDDMPRYVCEGTPIFSTATSLSDLTIESPPNELAAGEVGRGAQ 1680
Db 1621 QNRLOPKQHVSTFGDDMPRYVCEGTPIFSTATSLSDLTIESPPNELAAGEVGRGAQ 1680
Qy 1681 SGEFEKRDITPTEGRSTDEAOGGKTSVTTPELDDNKAEGDIIAECINAMPKKGSHKP 1740
Db 1681 SGEFEKRDITPTEGRSTDEAOGGKTSVTTPELDDNKAEGDIIAECINAMPKKGSHKP 1740
Qy 1741 FRVKIMDOVOOASASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNLN 1800
Db 1741 FRVKIMDOVOOASASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNLN 1800
Qy 1801 AERFSDNKSQKQNLKNNKDFNDKLPNNEEDRVGSGFAPDSPHHYTPIEGTPYCFGRND 1860
Db 1801 AERFSDNKSQKQNLKNNKDFNDKLPNNEEDRVGSGFAPDSPHHYTPIEGTPYCFGRND 1860
Qy 1861 SLSSLDFFDDDDVLSREKAEILRKAKENKESAKVTSHTELTSNQQSANKTQAIKQPINR 1920
Db 1861 SLSSLDFFDDDDVLSREKAEILRKAKENKESAKVTSHTELTSNQQSANKTQAIKQPINR 1920
Qy 1921 GQPKPILOKQSTFPQSSKDIIDRGAATDEKLQNFIAENTPVCFSHNSSLSSLSIDQENN 1980
Db 1921 GQPKPILOKQSTFPQSSKDIIDRGAATDEKLQNFIAENTPVCFSHNSSLSSLSIDQENN 1980
Qy 1981 NKNEPIKETPPDSQGSFPQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDDLLQ 2040
Db 1981 NKNEPIKETPPDSQGSFPQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDDLLQ 2040
Qy 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGILGDBDLTLDKDIOKRPDSEHGLSPDSNFD 2100
Db 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGILGDBDLTLDKDIOKRPDSEHGLSPDSNFD 2100
Qy 2101 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSLSLKSGISLGSPFHLTPDQEEKPFT 2160
Db 2101 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSLSLKSGISLGSPFHLTPDQEEKPFT 2160

Qy 2161 SNKGPRILKPGKSTLETYKKEIESESKGKGGKVKYKSLITGKVRNSSEISQMKQPLQAN 2220
Db 2161 SNKGPRILKPGKSTLETYKKEIESESKGKGGKVKYKSLITGKVRNSSEISQMKQPLQAN 2220
Qy 2221 MFSISGRPTMIHIGVRNSSSSTSPVSKGKPPKLTTPASKSPSEGOTATTSPGAKPSVKS 2280
Db 2221 MFSISGRPTMIHIGVRNSSSSTSPVSKGKPPKLTTPASKSPSEGOTATTSPGAKPSVKS 2280
Qy 2281 ELSPVARQTSQIGGSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNGISPN 2340
Db 2281 ELSPVARQTSQIGGSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNGISPN 2340
Qy 2341 KLSQLPRTSSPSTASTKSSGSKMYTSPGRMSOONLTQOTGLSKNASSIPRSESASG 2400
Db 2341 KLSQLPRTSSPSTASTKSSGSKMYTSPGRMSOONLTQOTGLSKNASSIPRSESASG 2400
Qy 2401 LQMNGNGANKVELSRMSTSKSSGESDRSERPVLVRQSTFIKEAPSPTLRRKLEESA 2460
Db 2401 LQMNGNGANKVELSRMSTSKSSGESDRSERPVLVRQSTFIKEAPSPTLRRKLEESA 2460
Qy 2461 SFESLSPSRPASPTRSQATPVLSPFLPDMSLSTHSSVQAGGWKLPNLSPTEYNDG 2520
Db 2461 SFESLSPSRPASPTRSQATPVLSPFLPDMSLSTHSSVQAGGWKLPNLSPTEYNDG 2520
Qy 2521 RPAKRHDIAARSHSESPSRLPINRSGTWKREHSHKSSSLPRVSTWRTGSSSILSASSES 2580
Db 2521 RPAKRHDIAARSHSESPSRLPINRSGTWKREHSHKSSSLPRVSTWRTGSSSILSASSES 2580
Qy 2581 SKKAXSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSPTNSTSTQTVSSGATNGAES 2640
Db 2581 SKKAXSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSPTNSTSTQTVSSGATNGAES 2640
Qy 2641 KTLIYQMAFAVAKTEDVWVRIEDCPINPRSGRSTPTGNTPPVIDSVSEKANPNIKDSKN 2700
Db 2641 KTLIYQMAFAVAKTEDVWVRIEDCPINPRSGRSTPTGNTPPVIDSVSEKANPNIKDSKN 2700
Qy 2701 QAKQNVGNGSVPMRTVGLNRLTSTFIQVADPQKGTETKPGQNNPVPVSETNESPVERT 2760
Db 2701 QAKQNVGNGSVPMRTVGLNRLTSTFIQVADPQKGTETKPGQNNPVPVSETNESPVERT 2760
Qy 2761 PFSSSSSSKHSSPSGTVAAARVTPFNYPNPSPRKSSADTSARPSQIPTVNNNTKKRDSKT 2820
Db 2761 PFSSSSSSKHSSPSGTVAAARVTPFNYPNPSPRKSSADTSARPSQIPTVNNNTKKRDSKT 2820
Qy 2821 DSTEESGTSQSPKXHSGLVTSV 2843
Db 2821 DSTEESGTSQSPKXHSGLVTSV 2843
RESULT 9
ABG71105
ID ABG71105 standard; protein; 2843 AA.
XX
AC ABG71105;
XX
DT 08-JAN-2003 (first entry)
XX Human adenomatous polyposis coli (APC) protein.
DE Human; adenomatous polyposis coli; APC; transgenic animal; colon cancer;
KW developmental abnormality; anticancer drug; beta-catenin;
KW Wnt/Wg signalling pathway.
XX Homo sapiens.
XX EP1243646-A2.
XX PD 25-SEP-2002.
XX 07-MAR-2002; 2002EP-00290571.
XX PR 19-MAR-2001; 2001US-0276483P.

XX PA (COUL) COUNCIL SCI & IND RES.
 XX PI Bhandari P, Shashidhara IS;
 XX PS WPI; 2002-742695/81.
 XX PT New transgenic Drosophila containing the human colon cancer gene
 PT Adenomatous Polyposis Coli is useful as an assay model to screen for new
 PT drugs, particularly against colon cancer.
 XX PS Claim 1; Page 19-20; 52pp; English.
 XX CC The invention describes a transgenic Drosophila whose genome comprises
 CC the full-length human colon cancer gene Adenomatous Polyposis Coli (APC)
 CC allowing regulated mis-expression of the APC gene resulting in
 CC developmental abnormalities. The transgenic flies are used to screen and
 CC validate efficacy of anticancer drugs, to identify new target proteins
 CC interacting with beta-catenin, genes which interact with human APC, to
 CC study the biochemical function of human APC and to identify additional
 CC components of the Drosophila Wnt/Mg signalling pathway. In particular the
 CC flies are used to screen potential drugs against colon cancer. This is
 CC the amino acid sequence of the human adenomatous polyposis coli (APC)
 CC protein DNA encoding which is incorporated in the transgenic flies
 XX Sequence 2843 AA;

Query Match 99.7%; Score 14533; DB 5; Length 2843;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAAASYDQLLKQVEALKVMSNLRQLEDNSNHLTKLETEASNKVELKQLOGSIEDEAM 60
 Db 1 MAAASYDQLLKQVEALKVMSNLRQLEDNSNHLTKLETEASNKVELKQLOGSIEDEAM 60

Qy 61 ASSGQDILLRLKELNLDSSNFFGVKLRKMSLRYSVSGREGSVSRSGCSVPVWGSFPR 120
 Db 61 ASSGQDILLRLKELNLDSSNFFGVKLRKMSLRYSVSGREGSVSRSGCSVPVWGSFPR 120

Qy 121 RGVNFSRGSTGYLEELKERSLLADLQKEEKDQWYQALQNLTKRIDSLPTNFSL 180
 Db 121 RGVNFSRGSTGYLEELKERSLLADLQKEEKDQWYQALQNLTKRIDSLPTNFSL 180

Qy 181 QDILTRQLEYEARQIRVAMEOGLGTCQDMEKAQRRIARIQIEKDILRIQLQSQT 240
 Db 181 QDILTRQLEYEARQIRVAMEOGLGTCQDMEKAQRRIARIQIEKDILRIQLQSQT 240

Qy 241 EARSQNKHETGSHDAERQNGGVGEINMATSNGQGSTTMDHETASVLSSTHSA 300
 Db 241 EARSQNKHETGSHDAERQNGGVGEINMATSNGQGSTTMDHETASVLSSTHSA 300

Qy 301 PRLTSLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
 Db 301 PRLTSLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360

Qy 361 HGNDSKVLGNRSGKEARASAAHNTIHSQDDKGRREIRVILHLEQIRAVCEFC 420
 Db 361 HGNDSKVLGNRSGKEARASAAHNTIHSQDDKGRREIRVILHLEQIRAVCEFC 420

Qy 421 WQEAHEPGMDQKNFMPAPVEHQICPAVCVLMKLSFDEEHRHAMNGLGQIAELLQ 480
 Db 421 WQEAHEPGMDQKNFMPAPVEHQICPAVCVLMKLSFDEEHRHAMNGLGQIAELLQ 480

Qy 481 VDCVMGLTNDHYSITLRRVAGVALTNLTFGVANKATILCSMKGCRVALVAQKSESDL 540
 Db 481 VDCVMGLTNDHYSITLRRVAGVALTNLTFGVANKATILCSMKGCRVALVAQKSESDL 540

Qy 541 QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEKVESTLSKLSALWNSAHC 600
 Db 541 QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEKVESTLSKLSALWNSAHC 600

Qy 601 ENKADICAVDGCALAFVGLTYRSQNTNLAIIBSGGGILRNVSLLIATNEDHQILRENN 660

Db 601 ENKADICAVDGCALAFVGLTYRSQNTNLAIIBSGGGILRNVSLLIATNEDHQILRENN 660

Qy 661 CLQTLQHLKSHSLTIVSNACGTLWNLARNPKDQEAALWDMGAVSMKLNLIHSHKXKIAM 720
 Db 661 CLQTLQHLKSHSLTIVSNACGTLWNLARNPKDQEAALWDMGAVSMKLNLIHSHKXKIAM 720

Qy 721 GSAALRNLMANRPAPKADANIMSPGSSLSLHVKKOKALEAELDQHLSETDNDNLS 780
 Db 721 GSAALRNLMANRPAPKADANIMSPGSSLSLHVKKOKALEAELDQHLSETDNDNLS 780

Qy 781 PKASHRSKQHRKQSLGYDYVFDNRHDDNRSDFNTGNMTVLSPYLNTTVLPSSSSRGS 840
 Db 781 PKASHRSKQHRKQSLGYDYVFDNRHDDNRSDFNTGNMTVLSPYLNTTVLPSSSSRGS 840

Qy 841 LDSRSRSDKSLERERGIGLGNYPATENGTSSKRGLOISTTAAQIAKWEVSAIHTS 900
 Db 841 LDSRSRSDKSLERERGIGLGNYPATENGTSSKRGLOISTTAAQIAKWEVSAIHTS 900

Qy 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSTNTYNTFTKSENSNRITCSPYAKLEYKRSS 960
 Db 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSTNTYNTFTKSENSNRITCSPYAKLEYKRSS 960

Qy 961 NDSLNSVSSNDGYKRGQMKPSTESYSEDDSKFCYGYQPADLAHKIHSANHMDNDGE 1020
 Db 961 NDSLNSVSSNDGYKRGQMKPSTESYSEDDSKFCYGYQPADLAHKIHSANHMDNDGE 1020

Qy 1021 LDTPIYSLKYDDEQLNSGRQSPSONERWARPKHIIIDEIKQSEQRQSRNOSTTYPVTE 1080
 Db 1021 LDTPIYSLKYDDEQLNSGRQSPSONERWARPKHIIIDEIKQSEQRQSRNOSTTYPVTE 1080

Qy 1081 STDDKHLKFPQHFQGCVCVPSYRSGANGSETNRVGNHGINQVNSOSLQCEDDYEDDKP 1140
 Db 1081 STDDKHLKFPQHFQGCVCVPSYRSGANGSETNRVGNHGINQVNSOSLQCEDDYEDDKP 1140

Qy 1141 TNYSERISEEHEEERPTNYSIKYNEERHVDQPIDYSLKYATDIPSSQKOSFSPFSKS 1200
 Db 1141 TNYSERISEEHEEERPTNYSIKYNEERHVDQPIDYSLKYATDIPSSQKOSFSPFSKS 1200

Qy 1201 SSGQSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260
 Db 1201 SSGQSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260

Qy 1261 TYCVEDTPICFSCSSLSLSAEDDEIGCNOTTQADDSANTLOIAETKGIKIGTRSAEDPV 1320
 Db 1261 TYCVEDTPICFSCSSLSLSAEDDEIGCNOTTQADDSANTLOIAETKGIKIGTRSAEDPV 1320

Qy 1321 SEVPAVQHPRTKSRRLQGSLSLSESAHKAIVFPSSGAKSPSKGATPKSPPEHYVQET 1380
 Db 1321 SEVPAVQHPRTKSRRLQGSLSLSESAHKAIVFPSSGAKSPSKGATPKSPPEHYVQET 1380

Qy 1381 PLMFSRCTSVSLDSFESRSIASVQSEPCSGVSGIIISPSDLPSDQCPMPSPSRKTPP 1440
 Db 1381 PLMFSRCTSVSLDSFESRSIASVQSEPCSGVSGIIISPSDLPSDQCPMPSPSRKTPP 1440

Qy 1441 PPPTAQTAKREVPKNAKPTAEKRESGPKQAANAAVQVQVLPDADTLHLHFAETESTPDGF 1500
 Db 1441 PPPTAQTAKREVPKNAKPTAEKRESGPKQAANAAVQVQVLPDADTLHLHFAETESTPDGF 1500

Qy 1501 SCSSSLSALSLEDPFIOKVELRIMPPVQNDNGNGETESEPQKESNENQKEAKIIDS 1560
 Db 1501 SCSSSLSALSLEDPFIOKVELRIMPPVQNDNGNGETESEPQKESNENQKEAKIIDS 1560

Qy 1561 KDLDDDDDDDIETLEBICIISAMPTKSRKKGKPAOTASKLPPPVARKPSQLPVYKLLPS 1620
 Db 1561 KDLDDDDDDDIETLEBICIISAMPTKSRKKGKPAOTASKLPPPVARKPSQLPVYKLLPS 1620

Qy 1621 QNRLOPKQHSFTPGDDMPRVYCVVEGTPINFSTATSLDITIESPPNVELAAGGVGGGAQ 1680
 Db 1621 QNRLOPKQHSFTPGDDMPRVYCVVEGTPINFSTATSLDITIESPPNVELAAGGVGGGAQ 1680

Qy 1681 SGFEKEDTPTTEGRSTDEAGGKTSSTVTIPELDDNKABEGDILAEICINSAMPKGSHPK 1740
 Db 1681 SGFEKEDTPTTEGRSTDEAGGKTSSTVTIPELDDNKABEGDILAEICINSAMPKGSHPK 1740

QY 1741 FRVKIMDQVQASASSAPNKNQDGGKKKPTSPVKPIPONTETRTVRKNADSKNNLN 1800
Db 1741 FRVKIMDQVQASASSAPNKNQDGGKKKPTSPVKPIPONTETRTVRKNADSKNNLN 1800
QY 1801 AERFVNKSKONLKNNSKDFNDKLPNNEDRVGSGFADSPSHYPTPIEGTYPVCFSRND 1860
Db 1801 AERFVNKSKONLKNNSKDFNDKLPNNEDRVGSGFADSPSHYPTPIEGTYPVCFSRND 1860
QY 1861 SLSLDFDDDDVDLSREKAEIRKAKENKESAKVTSHTLTSNQQSANKTOAIKOPINR 1920
Db 1861 SLSLDFDDDDVDLSREKAEIRKAKENKESAKVTSHTLTSNQQSANKTOAIKOPINR 1920
QY 1921 GQPKILQKSTFPQSSKDPDRGAATDEKLOFAIENTPVCFRSHNSLSLSIDIDENN 1980
Db 1921 GQPKILQKSTFPQSSKDPDRGAATDEKLOFAIENTPVCFRSHNSLSLSIDIDENN 1980
QY 1981 NKNEPIKETETPPDSQGEPSKQAGVAPKSFHVEDTPVCFRSHNSLSLSIDIDENN 2040
Db 1981 NKNEPIKETETPPDSQGEPSKQAGVAPKSFHVEDTPVCFRSHNSLSLSIDIDENN 2040
QY 2041 ECISAMPKPKKRLKDNKXHSFRNMGILGEDLTLDKDIORPDEHGLSPDSENF 2100
Db 2041 ECISAMPKPKKRLKDNKXHSFRNMGILGEDLTLDKDIORPDEHGLSPDSENF 2100
QY 2101 WKAIQEGANSIVSLHQAACLSROASDSDSLKSGISLGSPPHLPDQEEKPFT 2160
Db 2101 WKAIQEGANSIVSLHQAACLSROASDSDSLKSGISLGSPPHLPDQEEKPFT 2160
QY 2161 SNKGPRILKPKETLETKKIESBSKGIKGGKVKYKSLITGVKVSNSIEISQMKQPIQAN 2220
Db 2161 SNKGPRILKPKETLETKKIESBSKGIKGGKVKYKSLITGVKVSNSIEISQMKQPIQAN 2220
QY 2221 MPSISRGRTMIHIPVRNNSSTSPVSKGPKPLKTPASKSPSEGTATTSRPGAKPVSKS 2280
Db 2221 MPSISRGRTMIHIPVRNNSSTSPVSKGPKPLKTPASKSPSEGTATTSRPGAKPVSKS 2280
QY 2281 ELSPVARTSITGSSKAPSRSGRSDTSPRPAQPLSRPIQSPCRNSISPRNGISPPN 2340
Db 2281 ELSPVARTSITGSSKAPSRSGRSDTSPRPAQPLSRPIQSPCRNSISPRNGISPPN 2340
QY 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGQMSQONLTOTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGQMSQONLTOTGLSKNASSIPRSESASKG 2400
QY 2401 LNMNNGNGANKVLSRMSSTKSSGSDSRERVLVRQFTFIKEAPSLRRLKLEESA 2460
Db 2401 LNMNNGNGANKVLSRMSSTKSSGSDSRERVLVRQFTFIKEAPSLRRLKLEESA 2460
QY 2461 SPESLSPSRPASPTRSQAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPIEYNDG 2520
Db 2461 SPESLSPSRPASPTRSQAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPIEYNDG 2520
QY 2521 RPAKRHDIAHSBESPSRLPINRSGTWKREHSHSSLPVSTWTRTSSSSSILSASSES 2580
Db 2521 RPAKRHDIAHSBESPSRLPINRSGTWKREHSHSSLPVSTWTRTSSSSSILSASSES 2580
QY 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFPSTNTSOTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFPSTNTSOTVSSGATNGAES 2640
QY 2641 KTLIYQMAVASKTEVMTVRIEDCPNPNRSGRSPGTGTPVIDSVSEKANPNIKSDKN 2700
Db 2641 KTLIYQMAVASKTEVMTVRIEDCPNPNRSGRSPGTGTPVIDSVSEKANPNIKSDKN 2700
QY 2701 QAKQNVGNSVPMRTVGLNRLTSTFIQVDAPQKGTETKPGQNNVPVSETNESPIVERT 2760
Db 2701 QAKQNVGNSVPMRTVGLNRLTSTFIQVDAPQKGTETKPGQNNVPVSETNESPIVERT 2760
QY 2761 PFSSSSSSKHSPSGTVAARVTPFNPNPFRKSSADTSARSOIPTPVNNNTKRDST 2820
Db 2761 PFSSSSSSKHSPSGTVAARVTPFNPNPFRKSSADTSARSOIPTPVNNNTKRDST 2820

QY 2821 DSTESSTQSPKRHSGSYLTVSV 2843
Db 2821 DSTESSTQSPKRHSGSYLTVSV 2843
RESULT 10
AAW76821
ID AAW76821 standard; protein; 2973 AA.
XX AAW76821;
XX AC
XX XX
DT 25-JAN-1999 (first entry)
XX
DE Human APC protein.
XX
XX Tcf; beta-catenin; human; drug: familial adenomatous polyposis; FAP;
KW cancer; adenomatous polyposis coli; APC; neoplastic.
XX
OS Homo sapiens.
XX
XX MO9841631-A2.
XX
PD 24-SEP-1998.
XX
PF 20-MAR-1998; 98WO-US005506.
XX
PR 20-MAR-1997; 97US-00821355.
XX
XX (UYJO) UNIV JOHNS HOPKINS.
PA (UYUT-) RIJKSUNIV UTRECHT.
XX
XX Barker N, Clevers H, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
PI Vogelstein B;
PI
XX WPI; 1998-531569/45.
XX
XX
PT Intron-free DNA encoding Tcf-4 protein - useful for, e.g. identifying
PT drugs for treating FAP patients, or patients with increased risk of
XX developing cancer.
XX
PS Disclosure; Page 32-37; 58pp; English.
XX
XX This sequence represents a human APC protein which is used in a method of
CC identifying candidate drugs for use in familial adenomatous polyposis
CC (FAP) patients, or patients with increased risk of developing cancer. The
CC protein can also be used to determine the presence or absence in a cell
CC of wild type adenomatous polyposis coli (APC) gene or a downstream
CC protein in the APC transcription regulatory pathway. This method involves
CC introducing a Tcf-responsive reporter gene into the cell, and measuring
CC transcription of the reporter gene where a cell which supports active
CC downstream protein of the APC transcription regulatory pathway. The
CC protein can also be used in a method of diagnosing cancer in a sample
CC suspected of being neoplastic and for treating a patient with colorectal
CC cancer or other cancer associated with FAP, comprising administering to
CC the patient a polypeptide comprising a portion of the APC sequence,
CC comprising the beta-catenin binding site
XX
SQ Sequence 2973 AA;

Query Match 99.7%; Score 14533; DB 2; Length 2973;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAASYDQLLKQVEALKMENSLRQLEEDNSNHLTKLETSANNKVKLQKQSSIDEAM 60
Db 1 MAAASYDQLLKQVEALKMENSLRQLEEDNSNHLTKLETSANNKVKLQKQSSIDEAM 60
QY 61 ASSQIDILLERLKEINLDSNFPVGVKLRKWSLRYSRSGSVSSRSGECSVPVPMGSPFR 120
Db 61 ASSQIDILLERLKEINLDSNFPVGVKLRKWSLRYSRSGSVSSRSGECSVPVPMGSPFR 120
QY 121 RGVNRSRSTGYLEELERKERSULLADLKKEKEKDWYIAQLQNLTKRISLPLTENFSL 180

Db 121 RGVNGRESRGYLEEUEKERSLSLADLDEEKEKDWYAYQLOQLTKRIDSLPLTENFSL 180
Qy 181 QDILTRRQLEVEARQIRVAMEEQGLTQDMKEKAQRRIARIQOIEKDILIRIOLLOSQAT 240
Db 181 QDILTRRQLEVEARQIRVAMEEQGLTQDMKEKAQRRIARIQOIEKDILIRIOLLOSQAT 240
Qy 241 EABRSSQKHETGSHDAERONEGGVGEINWATSGNGCGSTTRMDHETASVLSSSSTHSA 300
Db 241 EABRSSQKHETGSHDAERONEGGVGEINWATSGNGCGSTTRMDHETASVLSSSSTHSA 300
Qy 301 PRRLTSHLGTKEVMYSLLSMLGTHDKDDMSRITLLAMSSSDSCISNROSGCLPILLOL 360
Db 301 PRRLTSHLGTKEVMYSLLSMLGTHDKDDMSRITLLAMSSSDSCISNROSGCLPILLOL 360
Qy 361 HGNDKDSVLGNRSGSKEAPARASAAALHNTIHSQDDKGRREIRVHLLEQIRAYCETC 420
Db 361 HGNDKDSVLGNRSGSKEAPARASAAALHNTIHSQDDKGRREIRVHLLEQIRAYCETC 420
Qy 421 WEOEAHEPGWODKNPMPAPVHOICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLO 480
Db 421 WEOEAHEPGWODKNPMPAPVHOICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLO 480
Qy 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTFDGVANKATLCSMKGCMRALVAQLKSESED 540
Db 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTFDGVANKATLCSMKGCMRALVAQLKSESED 540
Qy 541 QOVIASVLRNLWRADVNSKTLREVGSVKALMECALEVKKESTLKSVLNLSAHCT 600
Db 541 QOVIASVLRNLWRADVNSKTLREVGSVKALMECALEVKKESTLKSVLNLSAHCT 600
Qy 601 ENKADI CAVDGALAFVLGTLTYRSQNTLAIIBSGGILRNYSLLIATNEDHQILRENN 660
Db 601 ENKADI CAVDGALAFVLGTLTYRSQNTLAIIBSGGILRNYSLLIATNEDHQILRENN 660
Qy 661 CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDQBALMDMGAVSMKMLIHSKHKQIAM 720
Db 661 CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDQBALMDMGAVSMKMLIHSKHKQIAM 720
Qy 721 GSAAALRNLMANRPKADIANIMSGSSLP SLHVRKOKALEAELDAHLSETFDINILS 780
Db 721 GSAAALRNLMANRPKADIANIMSGSSLP SLHVRKOKALEAELDAHLSETFDINILS 780
Qy 781 PKASHRSKORHKQSLYGDYVDFNRRHDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSRGS 840
Db 781 PKASHRSKORHKQSLYGDYVDFNRRHDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSRGS 840
Qy 841 LQSSRSKORSLERERIGIGLVHPATENPGTSSKEGLOISTTAQIAKWEVSAIHTS 900
Db 841 LQSSRSKORSLERERIGIGLVHPATENPGTSSKEGLOISTTAQIAKWEVSAIHTS 900
Qy 901 QEDRSSGTTTELHCVTDERNALRRSSAAHTSNFTKSENSTRTCSMPYAKLEYKSS 960
Db 901 QEDRSSGTTTELHCVTDERNALRRSSAAHTSNFTKSENSTRTCSMPYAKLEYKSS 960
Qy 961 NDSLSVNSNDGCKRGOMKPSIESSEDESKFSGYQYPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLSVNSNDGCKRGOMKPSIESSEDESKFSGYQYPADLAHKIHSANHMDNDGE 1020
Qy 1021 LQTPINYSKYDEQLNSGRQSPQNERWARPKHIIEDBIKQSEQRQNRQSTTYPVYTE 1080
Db 1021 LQTPINYSKYDEQLNSGRQSPQNERWARPKHIIEDBIKQSEQRQNRQSTTYPVYTE 1080
Qy 1081 STDDKHLKOPHFQOECVPSYRSGANGSETNRYGNSHGINQNVQSILCOEDDYDDKP 1140
Db 1081 STDDKHLKOPHFQOECVPSYRSGANGSETNRYGNSHGINQNVQSILCOEDDYDDKP 1140
Qy 1141 TNYSERYSSEEHEERPTNYSIKNEEKRVADOPIDYSLKYATDIPSSOKQSPFSKS 1200
Db 1141 TNYSERYSSEEHEERPTNYSIKNEEKRVADOPIDYSLKYATDIPSSOKQSPFSKS 1200
Qy 1201 SSGQSKTEHMSSESTSTPSSNAKRONQLHPSSAQSGRSGQPOKAATCKVSSINQETIQ 1260

Db 1201 SSGQSKTEHMSSESTSTPSSNAKRONQLHPSSAQSGRSGQPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDDTPICFSRCSLSLSLSAARDEICNCTTOEADSANTLOIAEIKGKIGTSAEDPV 1320
Db 1261 TYCVEDDTPICFSRCSLSLSLSAARDEICNCTTOEADSANTLOIAEIKGKIGTSAEDPV 1320
Qy 1321 SEVPVAVSQHPRTKSRLOGSSLSSESAPHKAVEFPSPGAKSPSKGAQTPKSPPHYVQET 1380
Db 1321 SEVPVAVSQHPRTKSRLOGSSLSSESAPHKAVEFPSPGAKSPSKGAQTPKSPPHYVQET 1380
Qy 1381 PLMFSECTSVSLSLOSFSRSIASSVQSEPCSGMVGIIISPDLPDSQOTWPPRSKTPP 1440
Db 1381 PLMFSECTSVSLSLOSFSRSIASSVQSEPCSGMVGIIISPDLPDSQOTWPPRSKTPP 1440
Qy 1441 PPPOTAQTKREVPKNKAPTAEKRESGPKQAANVAQVQVLPDADTLHLHFATESTPDGF 1500
Db 1441 PPPOTAQTKREVPKNKAPTAEKRESGPKQAANVAQVQVLPDADTLHLHFATESTPDGF 1500
Qy 1501 SCSSLSLSALSDDEPFIQKDVLELRIMPPVQENDNGNETSEOPKESNENQKEAKTIDSE 1560
Db 1501 SCSSLSLSALSDDEPFIQKDVLELRIMPPVQENDNGNETSEOPKESNENQKEAKTIDSE 1560
Qy 1561 KDLDDSDDDDDIEILEECIISAMPTKSRKPKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KDLDDSDDDDDIEILEECIISAMPTKSRKPKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOQKHVSFTPGDDMPRVYCVGEGTPIINFSTATISLSDLTIESPPNELAAGBVGGAQ 1680
Db 1621 QNRLOQKHVSFTPGDDMPRVYCVGEGTPIINFSTATISLSDLTIESPPNELAAGBVGGAQ 1680
Qy 1681 SGEFEKDDTPTGERSTDEAGGKTSSVTIPELDDNKAEBGDIILAEICINAMPKGSHKP 1740
Db 1681 SGEFEKDDTPTGERSTDEAGGKTSSVTIPELDDNKAEBGDIILAEICINAMPKGSHKP 1740
Qy 1741 FRVKIMQVQOQASASSAPNQNOLDGKKKPTSPVKPIIPONTETRYTRVRKNADSKNLN 1800
Db 1741 FRVKIMQVQOQASASSAPNQNOLDGKKKPTSPVKPIIPONTETRYTRVRKNADSKNLN 1800
Qy 1801 AERFVSNDKSKKNLKNNSKDNFKLPNNEDVRGSPAFDPSHHYTPTEGTTCYCSRD 1860
Db 1801 AERFVSNDKSKKNLKNNSKDNFKLPNNEDVRGSPAFDPSHHYTPTEGTTCYCSRD 1860
Qy 1861 SLSLSDDDDDVLSREKAEELKAKENKESEAKVTSHTELTSNQOSANKTQAIKQPINR 1920
Db 1861 SLSLSDDDDDVLSREKAEELKAKENKESEAKVTSHTELTSNQOSANKTQAIKQPINR 1920
Qy 1921 GQPKPILOKOSTPQSSKDIIPDRGAATDEKLOQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Db 1921 GQPKPILOKOSTPQSSKDIIPDRGAATDEKLOQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Qy 1981 NKENEPIKETEPDPSQGEPSKPOASGYAPKSFHVEDTPVCFSRNSLSLSLSDIDQENN 2040
Db 1981 NKENEPIKETEPDPSQGEPSKPOASGYAPKSFHVEDTPVCFSRNSLSLSLSDIDQENN 2040
Qy 2041 ECSSAMPKKKXSRKLDNEKHSPRNMGGILGEDITLQNDKIDQRPDSBHLGSPDSENF 2100
Db 2041 ECSSAMPKKKXSRKLDNEKHSPRNMGGILGEDITLQNDKIDQRPDSBHLGSPDSENF 2100
Qy 2101 WKAIQEGANSIVSLSHQAACCLSRQASDSDSILSLKSGISLGSFPHLTPDQEKPF 2160
Db 2101 WKAIQEGANSIVSLSHQAACCLSRQASDSDSILSLKSGISLGSFPHLTPDQEKPF 2160
Qy 2161 SNKGPRIKPGEKSTLETKKIESEKIGKGVKYSKLTITGKVRNSSETISGQMKQPLQAN 2220
Db 2161 SNKGPRIKPGEKSTLETKKIESEKIGKGVKYSKLTITGKVRNSSETISGQMKQPLQAN 2220
Qy 2221 MPISIRGRTHIHPGVNRNSSTSPYKKGPPUKTPASKSPSEGOATTPSKGKPSVKS 2280
Db 2221 MPISIRGRTHIHPGVNRNSSTSPYKKGPPUKTPASKSPSEGOATTPSKGKPSVKS 2280
Qy 2281 ELSFVARQTSIQGSSKAPSRSGRSDTSPRPAQQLSRPIQSPGRNISPGRNGISPPN 2340
Db 2281 ELSFVARQTSIQGSSKAPSRSGRSDTSPRPAQQLSRPIQSPGRNISPGRNGISPPN 2340

Qy 2341 KLSOLPRTSPSTASTSSGKWSYTSPCROMSQOQLTKOTGLSKNASSIPESASKG 2400
 Db 2341 KLSOLPRTSPSTASTSSGKWSYTSPCROMSQOQLTKOTGLSKNASSIPESASKG 2400
 Qy 2401 LNMWNGGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKLEESA 2460
 Db 2401 LNMWNGGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKLEESA 2460
 Qy 2461 SFESLSPSPASTTRSOAQTPVLSPLPMSLSTHSSVQAGGWRKLPPLNLTIEVNDG 2520
 Db 2461 SFESLSPSPASTTRSOAQTPVLSPLPMSLSTHSSVQAGGWRKLPPLNLTIEVNDG 2520
 Qy 2521 RPAKHDIARSHSPSRPLNRSRGTWKREHSKSSSLPRVSTWRTTGGSSSSILSASSES 2580
 Db 2521 RPAKHDIARSHSPSRPLNRSRGTWKREHSKSSSLPRVSTWRTTGGSSSSILSASSES 2580
 Qy 2581 SEKAKEDEKHNISGTSKQKNOVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Db 2581 SEKAKEDEKHNISGTSKQKNOVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Qy 2641 KTLIYQMAPAVSKTEDVWRIEDCPINPRSGRPTGNTPPVIDSVSEKANPNIKOSKON 2700
 Db 2641 KTLIYQMAPAVSKTEDVWRIEDCPINPRSGRPTGNTPPVIDSVSEKANPNIKOSKON 2700
 Qy 2701 QAKQNVGNGSVPMRTVGLNRLTSFIQVDAPDQGTGTEIKQGNPNVPVSETNESPIVERT 2760
 Db 2701 QAKQNVGNGSVPMRTVGLNRLTSFIQVDAPDQGTGTEIKQGNPNVPVSETNESPIVERT 2760
 Qy 2761 PFSSSSSKHSRSGTVAARTVPNNVNSPKSSADSTSAEPSOIPTPVNNTKRDSKT 2820
 Db 2761 PFSSSSSKHSRSGTVAARTVPNNVNSPKSSADSTSAEPSOIPTPVNNTKRDSKT 2820
 Qy 2821 DSTESSGTQSPKRHSYGLVTSV 2843
 Db 2821 DSTESSGTQSPKRHSYGLVTSV 2843
 RESULT 11
 ID AAY72782
 AC AAY72782
 XX AAY72782;
 DT 31-MAY-2001 (first entry)
 DE Transcriptional activation protein #1 related to the invention.
 KW Human; Tcf-4 protein; transcriptional activation factor; beta-catenin;
 KW apoptosis; colorectal cancer; Familial Adenomatous Polyposis; FAP; APC;
 KW adenomatous polyposis coli; recombinant adenovirus; Ad-Mini-Me; therapy;
 KW GFP; green fluorescent protein; GFP/CAPC fusion protein; cytosstatic.
 XX Homo sapiens.
 OS WO200116167-A2.
 PN 08-MAR-2001.
 XX 29-AUG-2000; 2000WO-US023635.
 PF 01-SEP-1999; 99US-00388354.
 PR (UJO) UNIV JOHNS HOPKINS.
 PA Barker N, Clevers JC, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
 PI Vogelstein B, He T;
 DR WPI; 2001-226675/23.
 XX New fusion protein comprising an enzyme covalently linked to a portion of
 PT the adenomatous polyposis coli comprising its beta-catenin binding domain
 PT useful for inducing apoptosis or treating colorectal cancer.

XX Disclosure; Page 74-81; 83pp; English.
 PS The present sequence is a transcriptional activation protein related to
 XX the invention. The invention relates to human Tcf-4 proteins and their
 CC corresponding cDNA molecules which encodes transcriptional activation
 CC factors of human Tcf/Lef family. Human tcf-4 binds to beta-catenin and
 CC activates transcription in colorectal epithelial cells. Moreover it has
 CC been found that adenomatous polyposis coli (APC) regulates this
 CC transcriptional activation, by binding to beta-catenin. The invention
 CC also provides a recombinant adenovirus, Ad-Mini-Me ie., APC Minus its
 CC amino- and carboxyl- terminal Ends which expresses a fusion protein,
 CC green fluorescent protein (GFP)/CAPC containing GFP fused to the central
 CC third of APC which contains its beta-catenin binding domain useful for
 CC inducing apoptosis or treating colorectal cancer. These fusion proteins
 CC are useful for treating cancer, e.g. colorectal cancer, and other cancers
 CC associated with Familial Adenomatous Polyposis (FAP) or patients with
 CC increased risk of developing cancer. Human Tcf-4 cDNA provides an
 CC excellent system for screening agents for their ability to promote
 CC delivery, integration, hybridization, expression, replication or
 CC integration in cells or in an animal. It also provides methods for
 CC diagnosis cancer in a sample suspected of being neoplastic
 XX
 XX Sequence 2973 AA;
 Qy Query Match 99.7%; Score 14533; DB 4; Length 2973;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 MAAASVDQLLKQVEALKMENSRLROELEDNSNHLTKLETEASNKEVLKQLOGSIDEAM 60
 Db 1 MAAASVDQLLKQVEALKMENSRLROELEDNSNHLTKLETEASNKEVLKQLOGSIDEAM 60
 Qy 61 ASSGQIDLLERKELNLDSSNPFVKLRKMSLRSGSVSSRSGSCSPVPMGSPFR 120
 Db 61 ASSGQIDLLERKELNLDSSNPFVKLRKMSLRSGSVSSRSGSCSPVPMGSPFR 120
 Qy 121 RGFVNGSRSTGYLELEKERSLLADLDKEKEKDWYAAQNLTKRIDSPLTENFSL 180
 Db 121 RGFVNGSRSTGYLELEKERSLLADLDKEKEKDWYAAQNLTKRIDSPLTENFSL 180
 Qy 181 QDLTRRQLEYEARQIRVAMEBOLGTCQDMKEKAQRRIARIQOIEKDIIRQLLOSQAT 240
 Db 181 QDWTTRRQLEYEARQIRVAMEBOLGTCQDMKEKAQRRIARIQOIEKDIIRQLLOSQAT 240
 Qy 241 EAERSSONKHETGSHDAERQNEGQVGEINMATSNGQSGTTRMDHETASVLSSTHSA 300
 Db 241 EAERSSONKHETGSHDAERQNEGQVGEINMATSNGQSGTTRMDHETASVLSSTHSA 300
 Qy 301 PRRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
 Db 301 PRRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
 Qy 361 HGNDKDSVLLGNRSGSKEARASAAHLNTHSQPDQKGRREIRVLHLEQIRAYCETC 420
 Db 361 HGNDKDSVLLGNRSGSKEARASAAHLNTHSQPDQKGRREIRVLHLEQIRAYCETC 420
 Qy 421 WEWQEAHEPGMDQDKNPMAPVVEHQICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLQ 480
 Db 421 WEWQEAHEPGMDQDKNPMAPVVEHQICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLQ 480
 Qy 481 VDCEMYGLTNDHYSITLRRYAGMALNTHLPGDVANKATLCSMKGCMRALVAOKSSEDL 540
 Db 481 VDCEMYGLTNDHYSITLRRYAGMALNTHLPGDVANKATLCSMKGCMRALVAOKSSEDL 540
 Qy 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALNLSAHCT 600
 Db 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALNLSAHCT 600
 Qy 601 ENKADI CAVDGALAFVLGTITRYSQNTNLTAIIESGGGILRNVSLSIATNEDHQLRENN 660
 Db 601 ENKADI CAVDGALAFVLGTITRYSQNTNLTAIIESGGGILRNVSLSIATNEDHQLRENN 660

QY 661 CLOTLLOKLSHSLTIVSNACGTLWNLARNPKDQOEALWDMGAVSMLKNLIHSHKHMIAM 720
Db 661 CLOTLLOKLSHSLTIVSNACGTLWNLARNPKDQOEALWDMGAVSMLKNLIHSHKHMIAM 720
QY 721 GSAALANLWANRPKAKYDANIMSPGSSLPVLRKOKALEABEADAOHLSETFDNDNLS 780
Db 721 GSAALANLWANRPKAKYDANIMSPGSSLPVLRKOKALEABEADAOHLSETFDNDNLS 780
QY 781 PKSHRSKORHKSGLYDYYFDNRHDDNRSDFNTGNMTVLSPYLNNTVLPSSSSRGS 840
Db 781 PKSHRSKORHKSGLYDYYFDNRHDDNRSDFNTGNMTVLSPYLNNTVLPSSSSRGS 840
QY 841 LQSSRSKDLRERERGIGLGNYPHATENPGTSSKRGLOISTTAAQIAKYMEEVSAIHTS 900
Db 841 LQSSRSKDLRERERGIGLGNYPHATENPGTSSKRGLOISTTAAQIAKYMEEVSAIHTS 900
QY 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHNTYNTFTKSENRTCSMPYAKLEYKRSS 960
Db 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHNTYNTFTKSENRTCSMPYAKLEYKRSS 960
QY 961 NDSLSYSSNDGYGKRGOMKPSIESYSEDDDESFCYGOVPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLSYSSNDGYGKRGOMKPSIESYSEDDDESFCYGOVPADLAHKIHSANHMDNDGE 1020
QY 1021 LDPINYSKYDEQLNSGRQSPQNERWARPHHIIIDEIKQEQQRNSQSTTTPVYTE 1080
Db 1021 LDPINYSKYDEQLNSGRQSPQNERWARPHHIIIDEIKQEQQRNSQSTTTPVYTE 1080
QY 1081 STDDKHLKFPQHPGQOBCVSPYRSRGANGSETNRVGSNGINQVNSQSLCQEDDDDDKP 1140
Db 1081 STDDKHLKFPQHPGQOBCVSPYRSRGANGSETNRVGSNGINQVNSQSLCQEDDDDDKP 1140
QY 1141 TNSERYSEEEHEEBERPNYIKYNEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1200
Db 1141 TNSERYSEEEHEEBERPNYIKYNEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1200
QY 1201 SSGQSKTEHMSSESENTSTPSSNAQRQNLHPSSAQSPSGOPOKAATCKVSSINQETIQ 1260
Db 1201 SSGQSKTEHMSSESENTSTPSSNAQRQNLHPSSAQSPSGOPOKAATCKVSSINQETIQ 1260
QY 1261 TYCVETPTICFSCSSLSLSAEDIEGNOQTTEADSNLTQIAIEIKIGITRGAEDPV 1320
Db 1261 TYCVETPTICFSCSSLSLSAEDIEGNOQTTEADSNLTQIAIEIKIGITRGAEDPV 1320
QY 1321 SEVPAYSQHPRTKSRLOQSSLSSEBARHKAVEPPSGAKSPKSGAOTPKSPPEHYVQET 1380
Db 1321 SEVPAYSQHPRTKSRLOQSSLSSEBARHKAVEPPSGAKSPKSGAOTPKSPPEHYVQET 1380
QY 1381 PLMFSRCTSVSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPSPPGOTMPPSRSKTTP 1440
Db 1381 PLMFSRCTSVSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPSPPGOTMPPSRSKTTP 1440
QY 1441 PPQTQATKREVPKNAKPAETAKRESQPKAAVNAVQVLPDADTLHFAETSTPDGF 1500
Db 1441 PPQTQATKREVPKNAKPAETAKRESQPKAAVNAVQVLPDADTLHFAETSTPDGF 1500
QY 1501 SCSSLSLSALSLDEPFIQKQVELRIMPVQENDNGNETESEOQKESNQEKAETIDSE 1560
Db 1501 SCSSLSLSALSLDEPFIQKQVELRIMPVQENDNGNETESEOQKESNQEKAETIDSE 1560
QY 1561 KOLLDDSDDDDEIBECIISAMPTKSRKGGKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KOLLDDSDDDDEIBECIISAMPTKSRKGGKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
QY 1621 QNRLOPKHVSFTPGDDMPRVYCVETGPIINFSTATSLDLTIESPNNELAAEGVRGGAQ 1680
Db 1621 QNRLOPKHVSFTPGDDMPRVYCVETGPIINFSTATSLDLTIESPNNELAAEGVRGGAQ 1680
QY 1681 SGEFEKRDITPEGRSTDBAQGKTSSTVITPELDDNKAEBEGDILAEICINSAMPKSKHP 1740
Db 1681 SGEFEKRDITPEGRSTDBAQGKTSSTVITPELDDNKAEBEGDILAEICINSAMPKSKHP 1740
QY 1741 FRVVKIMDQVOQASASSAPNKNQLDGKKKPTSPVKRPIQNTYRTRVRKNADSKNNLN 1800

Db 1741 FRVVKIMDQVOQASASSAPNKNQLDGKKKPTSPVKRPIQNTYRTRVRKNADSKNNLN 1800
QY 1801 AERVPFNDKOSKQNLKXNSKDFNDKLPNNEDRVRSFAEDSPHHYTPIEGTYPYCFGRND 1860
Db 1801 AERVPFNDKOSKQNLKXNSKDFNDKLPNNEDRVRSFAEDSPHHYTPIEGTYPYCFGRND 1860
QY 1861 SLSLSDPDDDDDVLSREKAEIRKAKENKESFAKTSHTELTSNOOSANKTAQAKOPINR 1920
Db 1861 SLSLSDPDDDDDVLSREKAEIRKAKENKESFAKTSHTELTSNOOSANKTAQAKOPINR 1920
QY 1921 GQPPILQKQSTFFQSSKDIPIRGAATDEKLQFAIENTPVCFSHNSLSLSLSDIDENN 1980
Db 1921 GQPPILQKQSTFFQSSKDIPIRGAATDEKLQFAIENTPVCFSHNSLSLSLSDIDENN 1980
QY 1981 NKENPIKETEPPOSQEPSPKQASGVAPKSFHVEDTPVCFSRNSSLSLSISDEDDLLQ 2040
Db 1981 NKENPIKETEPPOSQEPSPKQASGVAPKSFHVEDTPVCFSRNSSLSLSISDEDDLLQ 2040
QY 2041 ECISAMPKPKKPRKLGDNKXSPRNMGILGEBDLTLDLKDITORPDSHGLSPDSENF 2100
Db 2041 ECISAMPKPKKPRKLGDNKXSPRNMGILGEBDLTLDLKDITORPDSHGLSPDSENF 2100
QY 2101 WKAIQEGANSIVSSLHQAAAAACLSRQASDSDSIISLKSIGISLGSPFHILTPDQEKPF 2160
Db 2101 WKAIQEGANSIVSSLHQAAAAACLSRQASDSDSIISLKSIGISLGSPFHILTPDQEKPF 2160
QY 2161 SNKGPRILKPEKSTLTETKIESESGIKGKKVYKSLITGKVRNSSEISGOMKQLOAN 2220
Db 2161 SNKGPRILKPEKSTLTETKIESESGIKGKKVYKSLITGKVRNSSEISGOMKQLOAN 2220
QY 2221 MPSISRGKTIHIPGVNRNSSSTSPVSKGPPKLPASKSPSEGTATTPRGAKEPVKS 2280
Db 2221 MPSISRGKTIHIPGVNRNSSSTSPVSKGPPKLPASKSPSEGTATTPRGAKEPVKS 2280
QY 2281 ELSVPARQTOIGGSSKAPSRSGRSTSPRPAQOPLSRPIQSPGRNSISPGRNGISPPN 2340
Db 2281 ELSVPARQTOIGGSSKAPSRSGRSTSPRPAQOPLSRPIQSPGRNSISPGRNGISPPN 2340
QY 2341 KLSQLPRTSSPSTASTKSSGSGKVSYSYTPGRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSSPSTASTKSSGSGKVSYSYTPGRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
QY 2401 LNQMANGANGKVELSRMSSTKSSGSESDRSRPRVLRQSTFIKEAPSTLRKLEESA 2460
Db 2401 LNQMANGANGKVELSRMSSTKSSGSESDRSRPRVLRQSTFIKEAPSTLRKLEESA 2460
QY 2461 SFESLSPSSRPASPTRSQOQTPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
Db 2461 SFESLSPSSRPASPTRSQOQTPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
QY 2521 RPAKRHDIAHSHSESRLPINSRGTWKEHSHKSSSLPRVSTWRTGSSSSILSASSSS 2580
Db 2521 RPAKRHDIAHSHSESRLPINSRGTWKEHSHKSSSLPRVSTWRTGSSSSILSASSSS 2580
QY 2581 SEKAKSEDEKHVANSISGTQKSENOVSAGKTWRKIKENEFSPNTSQTSSGATNGASS 2640
Db 2581 SEKAKSEDEKHVANSISGTQKSENOVSAGKTWRKIKENEFSPNTSQTSSGATNGASS 2640
QY 2641 KTLIYQAPAVKTEDVWVRIEDCPIINPRSGSPGTNTPPVIDSVESEKANPNIKDSKN 2700
Db 2641 KTLIYQAPAVKTEDVWVRIEDCPIINPRSGSPGTNTPPVIDSVESEKANPNIKDSKN 2700
QY 2701 QAKQNVGSGVPMRTVGLNRLTSFIQVADPOKQTEIKPGQNNPVVSETNESPIVERT 2760
Db 2701 QAKQNVGSGVPMRTVGLNRLTSFIQVADPOKQTEIKPGQNNPVVSETNESPIVERT 2760
QY 2761 PFSSSSSKHSSPSGTVAAARVTPFNTPNPRKSSADSTGARPSSQIPTPVNNNTKKRDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAAARVTPFNTPNPRKSSADSTGARPSSQIPTPVNNNTKKRDSKT 2820
QY 2821 DSTESSGTQSPKEHSGSYLVTSV 2843

Db2821 DSTESSGTQSPKRGHSGSYLVTSV 2843

RESULT 12

AAAY70304

IDAAAY70304 standard; protein; 2973 AA.

XXAAAY70304;

AC06-JUN-2000 (first entry)

DTProtein used in cancer diagnosis associated with APC/beta-catenin.

DEHuman Tcf-4E protein; transcription factor; beta-catenin;

XXAdenomatous polyposis coli; APC; transcriptional activation;

XXTcf responsive reporter gene; APC transcriptional regulatory pathway;

XXfamilial adenomatous polyposis; FAP; cancer; colorectal; thyroid; brain;

XXmedulloblastoma; breast; head; neck; desmoid tumour; osteoma; cytostatic.

XXHomo sapiens.

OSWO200011195-A1.

XX02-MAR-2000.

PD20-AUG-1999; 99WO-US018774.

XX20-AUG-1998; 98US-00136605.

PR20-AUG-1998; 98US-00136605.

XX(UTJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

PAHe T, Vogelstein B, Kinzler KW;

XXWPI; 2000-237657/20.

XXDetermining wild-type adenomatous polyposis coli protein for diagnosing

PTcancer comprises introducing a Tcf responsive reporter gene having

PTupstream sequences of c-MYC into a cell.

XXDisclosure; Page 58-65; 70pp; English.

XXThe present sequence is a protein involved in cancer diagnosis associated

XXwith APC or beta-catenin mutations. Human Tcf-4E protein expressed in

XXcolorectal epithelium transactivates transcription when associated with

XXbeta-catenin. Adenomatous polyposis coli (APC) regulates this

XXtranscriptional activation, at least in part by binding to beta-catenin.

XXDetermining wild-type APC protein for diagnosing cancer comprises

XXintroducing a Tcf responsive reporter gene having upstream sequences of c

XX-MYC into a cell and measuring transcription of the reporter gene. The

XXcandidate drug identified is useful for treating familial adenomatous

XXpolyposis patients with APC or beta-catenin mutations and patients with

XXincreased risk of developing cancers such as colorectal, thyroid, brain,

XXmedulloblastoma, desmoid tumour, osteoma, breast, head and neck

XXSequence 2973 AA;

SQQuery Match 99.7%; Score 14526; DB 3; Length 2973;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 2835; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY1 MAAASYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKREVLKQLOGSIEDAM 60

DB1 MAAASYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKREVLKQLOGSIEDAM 60

QY61 ASSGQDILLERLKEMLNDSNFPVGLRKMSLRSYSGREGSVSRSGSCSPVPMGSPPR 120

DB61 ASSGQDILLERLKEMLNDSNFPVGLRKMSLRSYSGREGSVSRSGSCSPVPMGSPPR 120

QY121 RGFVNGSRSTGYLEELKERSLLADLDKEEKEDWYVAQLQNLTKRIDSPLTENFSL 180

DB121 RGFVNGSRSTGYLEELKERSLLADLDKEEKEDWYVAQLQNLTKRIDSPLTENFSL 180

QY181 QDLTRRQLEYEARQIRVAMEEQGTCDQMKRAQRRIARIQOIEKDILRIQLQSQT 240

Db181 QDLTRRQLEYEARQIRVAMEEQGTCDQMKRAQRRIARIQOIEKDILRIQLQSQT 240

QY241 EAERSQNKHETGSHDAERQNEGQGVGINMATSGOGSTTRMDHETASVLSSTHSA 300

DB241 EAERSQNKHETGSHDAERQNEGQGVGINMATSGOGSTTRMDHETASVLSSTHSA 300

QY301 PRRLTSHLGTQVEMVYSLLSMLGTHDKDMSTILLAMSSQSDSCISMRSQSGCLPLLIQL 360

DB301 PRRLTSHLGTQVEMVYSLLSMLGTHDKDMSTILLAMSSQSDSCISMRSQSGCLPLLIQL 360

QY361 HGNKDQSVLLGNSRGSKEARARASAAALHNIHQSOPDKRGREIRIVLHLEQIRAYCETC 420

DB361 HGNKDQSVLLGNSRGSKEARARASAAALHNIHQSOPDKRGREIRIVLHLEQIRAYCETC 420

QY421 WEMQEAHEFGMDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMELGGLOAIAELIQ 480

DB421 WEMQEAHEFGMDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMELGGLOAIAELIQ 480

QY481 VCEMYGLTNDHYSITLRYAGMALNTLTFGVANKATLCSMKGCWALVAQKSESEDL 540

DB481 VCEMYGLTNDHYSITLRYAGMALNTLTFGVANKATLCSMKGCWALVAQKSESEDL 540

QY541 QQVIASVLRNLNLSWRADVNSKKTIREVGSVKALMECALEVYKESSTLKSVLNLSAHCT 600

DB541 QQVIASVLRNLNLSWRADVNSKKTIREVGSVKALMECALEVYKESSTLKSVLNLSAHCT 600

QY601 ENKADI CAVDGALAFVGLTLYRSQNTLAIIESGGIILRNVSLLIATNEDHQILRENN 660

DB601 ENKADI CAVDGALAFVGLTLYRSQNTLAIIESGGIILRNVSLLIATNEDHQILRENN 660

QY661 CLQTLQHLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSMKLNLIHSHKHMIAM 720

DB661 CLQTLQHLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSMKLNLIHSHKHMIAM 720

QY721 GSAALRNLMANPAKYKDANIMSPGSLPSLHVRKQKALEAEALDAHLSEFDNLNLS 780

DB721 GSAALRNLMANPAKYKDANIMSPGSLPSLHVRKQKALEAEALDAHLSEFDNLNLS 780

QY781 PKASHRSKORHKOSLYGDYVDFDNRDNDNRDNFNNTGNMTVLSPLYNTTVLPSSSSSRGS 840

DB781 PKASHRSKORHKOSLYGDYVDFDNRDNDNRDNFNNTGNMTVLSPLYNTTVLPSSSSSRGS 840

QY841 LDSRSSEKDSLRERIGLGNVHPATENPGTSKRGLOISTTAAQIAKWEVSAIHTS 900

DB841 LDSRSSEKDSLRERIGLGNVHPATENPGTSKRGLOISTTAAQIAKWEVSAIHTS 900

QY901 QEDRSSGTTTELHCVTDERNALRRSAAHSTNYFTKSENNRCTCSMPYAKLEYKRSS 960

DB901 QEDRSSGTTTELHCVTDERNALRRSAAHSTNYFTKSENNRCTCSMPYAKLEYKRSS 960

QY961 NDSLNSVSSNDGYGKRGQMKPSIESYSEDDSKFCSYGQVPADLAHKIHSANHMDNDGE 1020

DB961 NDSLNSVSSNDGYGKRGQMKPSIESYSEDDSKFCSYGQVPADLAHKIHSANHMDNDGE 1020

QY1021 LDTPINYSILKYSDHOLNSGRQSPQNERWARPKHIIIEDEIKQSEQRNRQSTTYPVYTE 1080

DB1021 LDTPINYSILKYSDHOLNSGRQSPQNERWARPKHIIIEDEIKQSEQRNRQSTTYPVYTE 1080

QY1081 STDDKHLKFPQHFQOQCVSPYRSRGANGSETNRVSGNHGINQNVQSGLCOEDDYEDDKP 1140

DB1081 STDDKHLKFPQHFQOQCVSPYRSRGANGSETNRVSGNHGINQNVQSGLCOEDDYEDDKP 1140

QY1141 TNYSERYSBEEHQBEEBEPNTNYSIKYNEBKRVDPQIDYSLKYATDTPSSQKQSFSPKS 1200

DB1141 TNYSERYSBEEHQBEEBEPNTNYSIKYNEBKRVDPQIDYSLKYATDTPSSQKQSFSPKS 1200

QY1201 SSGOSSKTEHMSSESSENTSTPSSNAKQONOLHPSAQRSGQPKAATCKVSSINQETIQ 1260

DB1201 SSGOSSKTEHMSSESSENTSTPSSNAKQONOLHPSAQRSGQPKAATCKVSSINQETIQ 1260

QY1261 TYCVEDTPICFSRCSLSLSSAEDETGQNTTQOASANTLQIAETKGIKGTSAEDPV 1320

Db 1261 TYCVEDTPICFRCSSLSLSAEDIEGNCOTTQADSAANTLOIAEIKKIGTRSAEDPV 1320
 Qy 1321 SEVPAVSHPRTKSRLOGSSLSSESARHKAVEPPSGAKSPSKGAOTPKSPHYYQET 1380
 Db 1321 SEVPAVSHPRTKSRLOGSSLSSESARHKAVEPPSGAKSPSKGAOTPKSPHYYQET 1380
 Qy 1381 PLMFSTRCTSVSSLSDFSRSIASSVQSEPCSGMWGIIISPSDLDPDSQGTMPSPRSKTPP 1440
 Db 1381 PLMFSTRCTSVSSLSDFSRSIASSVQSEPCSGMWGIIISPSDLDPDSQGTMPSPRSKTPP 1440
 Qy 1441 PPPOTATKEVPKPKAPTAKEKREGSKQAANVAORVQVLPDADTLLHFAESTPDGF 1500
 Db 1441 PPPOTATKEVPKPKAPTAKEKREGSKQAANVAORVQVLPDADTLLHFAESTPDGF 1500
 Qy 1501 SCSSLSALSILDEFFIOKVELRMPVPQENDNGNETESQPKESNENQKEAEKTIIDSE 1560
 Db 1501 SCSSLSALSILDEFFIOKVELRMPVPQENDNGNETESQPKESNENQKEAEKTIIDSE 1560
 Qy 1561 KOLLDDDDDDIELEECIIISAMPTKSRKKGKPAQATASKLPPPVARKSOLPVYKLLPS 1620
 Db 1561 KOLLDDDDDDIELEECIIISAMPTKSRKKGKPAQATASKLPPPVARKSOLPVYKLLPS 1620
 Qy 1621 QNRLOPKHVSFTPGDDMPVYCEVETPIINFSTATSLDITIESPPNELAAGGVGGGAQ 1680
 Db 1621 QNRLOPKHVSFTPGDDMPVYCEVETPIINFSTATSLDITIESPPNELAAGGVGGGAQ 1680
 Qy 1681 SGEFEKEDTTPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAECSAMPKGSHP 1740
 Db 1681 SGEFEKEDTTPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAECSAMPKGSHP 1740
 Qy 1741 FRVKKIMDQVQOASASSAENKQLODKKKPTSPVKPIPONTTEYTRVRKKAADSKNIN 1800
 Db 1741 FRVKKIMDQVQOASASSAENKQLODKKKPTSPVKPIPONTTEYTRVRKKAADSKNIN 1800
 Qy 1801 AERFSDNCKSKQNLKNSKDFNDKLPNNEDVRGSAFDPSPHYTPIEGTPYCFSRND 1860
 Db 1801 AERFSDNCKSKQNLKNSKDFNDKLPNNEDVRGSAFDPSPHYTPIEGTPYCFSRND 1860
 Qy 1861 SLSSLOPDDDDVLSREKAEKRAKENKESEAKVTSHTLTNSQOQANKTQIAKQPINR 1920
 Db 1861 SLSSLOPDDDDVLSREKAEKRAKENKESEAKVTSHTLTNSQOQANKTQIAKQPINR 1920
 Qy 1921 GQPKPILQKSTFPQSSKDIPOKGAATDEKLQNFALIENTPVCFSHNSLSLSDIDQENN 1980
 Db 1921 GQPKPILQKSTFPQSSKDIPOKGAATDEKLQNFALIENTPVCFSHNSLSLSDIDQENN 1980
 Qy 1981 KENEDIKETEPDQSGEFPKQASGYAPKSFHVEDTPVCFSRNSLSLSDIDSEDDLLQ 2040
 Db 1981 KENEDIKETEPDQSGEFPKQASGYAPKSFHVEDTPVCFSRNSLSLSDIDSEDDLLQ 2040
 Qy 2041 ECISAMPKKKPSRLKGNKNSPRNMGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKKKPSRLKGNKNSPRNMGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
 Qy 2101 WKAIQEGANSIVSLHQAACLSRQASDSDSILSLKSGISLGSPFHLTPDQEKPT 2160
 Db 2101 WKAIQEGANSIVSLHQAACLSRQASDSDSILSLKSGISLGSPFHLTPDQEKPT 2160
 Qy 2161 SNKGPRILKPEKSTLETKEIESEKIGKGVKYSLTGKVRNSNISGOMQPLOAN 2220
 Db 2161 SNKGPRILKPEKSTLETKEIESEKIGKGVKYSLTGKVRNSNISGOMQPLOAN 2220
 Qy 2221 MPSISGRRTMIHPGVNRSSSTSPVSKGPPKTPASKSPSEGTATTSRGAKPSVKS 2280
 Db 2221 MPSISGRRTMIHPGVNRSSSTSPVSKGPPKTPASKSPSEGTATTSRGAKPSVKS 2280
 Qy 2281 ELSPVARQTSQIGGSKAPSRSGSDSTSPRAQOPLSRPTQSPGRNISPRNGISPPN 2340
 Db 2281 ELSPVARQTSQIGGSKAPSRSGSDSTSPRAQOPLSRPTQSPGRNISPRNGISPPN 2340
 Qy 2341 KLSQLPRTSPSTAKTSKSGSKMSYTSFGRQMSQONLTQOTGLSKNASSIPRESASKG 2400
 Db 2341 KLSQLPRTSPSTAKTSKSGSKMSYTSFGRQMSQONLTQOTGLSKNASSIPRESASKG 2400

Qy 2401 LNMONGANGKVKVLSRMSSTKSGSSEDSRSPVLVRQSTFFIKEAPSPTLRRKLEESA 2460
 Db 2401 LNMONGANGKVKVLSRMSSTKSGSSEDSRSPVLVRQSTFFIKEAPSPTLRRKLEESA 2460
 Qy 2461 SFESLSPSRSPASPTRSOAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Db 2461 SFESLSPSRSPASPTRSOAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Qy 2521 RPAKHDIARSHSPSPRLPINRSGTWKREHSKSSSLPRVSTWRTGSSSSSILSASSES 2580
 Db 2521 RPAKHDIARSHSPSPRLPINRSGTWKREHSKSSSLPRVSTWRTGSSSSSILSASSES 2580
 Qy 2581 SEKAKSEDEKHVNSTSGTKQSKENOVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Db 2581 SEKAKSEDEKHVNSTSGTKQSKENOVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Qy 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKN 2700
 Qy 2701 QAKQNVGNSVPMETVGLNRLTSTFIOVDAPDQKGTETIKPGQNNPVVSETNESPIVERT 2760
 Db 2701 QAKQNVGNSVPMETVGLNRLTSTFIOVDAPDQKGTETIKPGQNNPVVSETNESPIVERT 2760
 Qy 2761 PFSSSSSKHSSPSGTVAARVTPENYNPSPRKSADSTARSPOIPTPVNNTKKRDSKT 2820
 Db 2761 PFSSSSSKHSSPSGTVAARVTPENYNPSPRKSADSTARSPOIPTPVNNTKKRDSKT 2820
 Qy 2821 DSTESSGTQSPRHSGSYLVTSV 2843
 Db 2821 DSTESSGTQSPRHSGSYLVTSV 2843

RESULT 13

AAR58634
 ID AAR58634 standard; protein; 2843 AA.
 XX AC AAR58634;
 XX AC
 DT 25-MAR-2003 (revised)
 DT 21-JUN-1995 (first entry)
 XX XX Adenomatous polyposis coli protein (APC).
 DE XX
 XX Adenomatous polyposis coli protein gene; familial adenomatous polyposis;
 KW Adenomatous polyposis coli protein gene; familial adenomatous polyposis;
 KW colorectal tumor; adenoma.
 XX OS Homo sapiens.
 OS
 PN W09421814-A1.
 XX
 PD 29-SEP-1994.
 XX
 PF 21-MAR-1994; 94WO-US002987.
 XX
 PR 19-MAR-1993; 93US-00034850.
 XX
 XX (UYJO) UNIV JOHNS HOPKINS.
 FA Hall DE, Johnson KA, Kinzler KW, Vogelstein B;
 XX WPI; 1994-317033/39.
 DR N-PSDB; AAQ70633.
 XX
 PT Antibodies to adenomatous polyposis coli protein - are used for detecting
 PT mutations in the APC gene for predicting pre-disposition to cancer,
 PT partic. colon cancer.
 XX
 PS Claim 1; Page 46; 81pp; English.
 XX
 CC Antibodies prepared to this protein are used for detecting mutations in
 CC the APC gene for predicting pre-disposition to cancer, particularly colon

CC	cancer. The Abs can detect germ line or somatic mutations indicating a
CC	predisposition to colon cancer and possibly gastric, oesophageal,
CC	pancreatic or small cell lung cancers. (Updated on 25-MAR-2003 to correct
CC	PN field.)
XX	
SQ	Sequence 2843 AB;
	Query Match 99.6%; Score 14521; DB 2; Length 2843;
	Best Local Similarity 99.6%; Pred. No. 0;
	Matches 2833; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
Qy	1 MAASYPOLLKQVLEALWENSLFQELDENSHTLKLETEASNMKEVYLKOLQSGIDEAM 60
Db	
Qy	1 MAASYPOLLKQVLEALWENSLFQELDENSHTLKLETEASNMKEVYLKOLQSGIDEAI 60
Db	
Qy	61 ASSQOIDLRLKELNLDSSNFPVGLRSKMSLSYSGRSVSSRSGECSPVPMGSPFR 120
Db	
Qy	121 RGFVNGSRSTGYLEELKESRLLIADLDKEEKKWYLAQNLTKRIIDSLPTENFSL 180
Db	
Qy	121 RGFVNGSRSTGYLEELKESRLLIADLDKEEKKWYLAQNLTKRIIDSLPTENFSL 180
Db	
Qy	181 QTDLTRRQLEYEARQIRVAMEEQLGTCQDMKRAQRRIARIQOIEKDILRIQLQSOAT 240
Db	
Qy	181 QTDLTRRQLEYEARQIRVAMEEQLGTCQDMKRAQRRIARIQOIEKDILRIQLQSOAT 240
Db	
Qy	241 EABRSQNKHETGSHDABRQEGVGEINMATSNGGSTRMDHETASVLSSTHSA 300
Db	
Qy	241 EABRSQNKHETGSHDABRQEGVGEINMATSNGGSTRMDHETASVLSSTHSA 300
Db	
Qy	301 PRLTSHLGTQVEMVYLSLMLGTHDXDDMRTLLAMSSODSCISMRQSCCLPLLIQLL 360
Db	
Qy	301 PRLTSHLGTQVEMVYLSLMLGTHDXDDMRTLLAMSSODSCISMRQSCCLPLLIQLL 360
Db	
Qy	361 HGNDXDSVLLGNSRGSKEARAPASAAALNIIHSQPDQKGRREIRVLHLLQIRAYCETC 420
Db	
Qy	361 HGNDXDSVLLGNSRGSKEARAPASAAALNIIHSQPDQKGRREIRVLHLLQIRAYCETC 420
Db	
Qy	421 MEWQSAHEFGMDQXNPAPVHEHOICPAVCVLMKLSFDEHRHAMNELGGIQAIAELLQ 480
Db	
Qy	421 MEWQSAHEFGMDQXNPAPVHEHOICPAVCVLMKLSFDEHRHAMNELGGIQAIAELLQ 480
Db	
Qy	481 VDCEMYGLTNDHYSITLARYAGMALTNLTFGDVANKATLCMSKGCWALVAQLKSESDL 540
Db	
Qy	481 VDCEMYGLTNDHYSITLARYAGMALTNLTFGDVANKATLCMSKGCWALVAQLKSESDL 540
Db	
Qy	541 QOVIASVLRNLNWRADVNSKTLREVGSVKALMECALEVYKKESTLKSVLNLSAHSCT 600
Db	
Qy	541 QOVIASVLRNLNWRADVNSKTLREVGSVKALMECALEVYKKESTLKSVLNLSAHSCT 600
Db	
Qy	601 ENKADICAVDGGALAFVGLTVRSQNTLAIIESGGIILRVNSSLIAINEDHRQILRENN 660
Db	
Qy	601 ENKADICAVDGGALAFVGLTVRSQNTLAIIESGGIILRVNSSLIAINEDHRQILRENN 660
Db	
Qy	661 CLQTLQLHLSHLTIVSNACGTLNLSARNPKDQBALMDGAVSMKLNLIHSKHKMTAM 720
Db	
Qy	661 CLQTLQLHLSHLTIVSNACGTLNLSARNPKDQBALMDGAVSMKLNLIHSKHKMTAM 720
Db	
Qy	721 GSAALRNLMANRPARYKQANIMSPGSLPSLHVRKQKALELDAOHLSETFDNIDNLS 780
Db	
Qy	721 GSAALRNLMANRPARYKQANIMSPGSLPSLHVRKQKALELDAOHLSETFDNIDNLS 780
Db	
Qy	781 PKASHRSKQPHKQSLGYDVFDTNRDNRDNFNFGNMVLSPLYNTTVLPSSSSSRGS 840
Db	
Qy	781 PKASHRSKQPHKQSLGYDVFDTNRDNRDNFNFGNMVLSPLYNTTVLPSSSSSRGS 840
Db	
Qy	841 LDSRSREKDSLRERIGLGNVHPATENPGTSKKEGLQISTTAAQIAKMEVEVAIHTS 900
Db	
Qy	841 LDSRSREKDSLRERIGLGNVHPATENPGTSKKEGLQISTTAAQIAKMEVEVAIHTS 900
Db	
Qy	901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHNTYNTFTKSENSNRTCSPYAKLEYKRSS 960
Db	

Db	901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHNTYNTFTKSENSNRTCSPYAKLEYKRSS 960
Qy	961 NDSLSVSNNDGYGRGQMKPISIESYEDDSDSKFCYSQYPAHLAKHSHANHMNDNDE 1020
Db	
Qy	961 NDSLSVSNNDGYGRGQMKPISIESYEDDSDSKFCYSQYPAHLAKHSHANHMNDNDE 1020
Db	
Qy	1021 LQTPNYSILKYDEOLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Db	
Qy	1021 LQTPNYSILKYDEOLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Db	
Qy	1081 STDDKHLKFPQHFQOECVSPYRSRGANGSETRVGSNHNINQNVQSILCOBDDYEDDKP 1140
Db	
Qy	1081 STDDKHLKFPQHFQOECVSPYRSRGANGSETRVGSNHNINQNVQSILCOBDDYEDDKP 1140
Db	
Qy	1141 TNYSEYSEEEHREERTNYSIKYNEKRVHVDOPIDYSLKYATDIPSSQKQSFSSKS 1200
Db	
Qy	1141 TNYSEYSEEEHREERTNYSIKYNEKRVHVDOPIDYSLKYATDIPSSQKQSFSSKS 1200
Db	
Qy	1201 SSQSSKTEHMSSESTSTPSSNAKQNLHPSSAQSSGQPKAAATCKVSIHQETIQ 1260
Db	
Qy	1201 SSQSSKTEHMSSESTSTPSSNAKQNLHPSSAQSSGQPKAAATCKVSIHQETIQ 1260
Db	
Qy	1261 TYCVETPTICFRCSSLSLSAEDIGCNQTOEADSAANTIQAIEIKGIGTRSAEDPV 1320
Db	
Qy	1261 TYCVETPTICFRCSSLSLSAEDIGCNQTOEADSAANTIQAIEIKGIGTRSAEDPV 1320
Db	
Qy	1321 SEVPAYSQHPRTKSSSLQSSLSSESARHKAVEFFPSGAKSPKSGAQTPKSPPEHVYQET 1380
Db	
Qy	1321 SEVPAYSQHPRTKSSSLQSSLSSESARHKAVEFFPSGAKSPKSGAQTPKSPPEHVYQET 1380
Db	
Qy	1381 PLMFSRCTSVSSLDSPESRSIASVQSEPCSGMVSCIISPSDLPDPSGOTMPPSRSKTTP 1440
Db	
Qy	1381 PLMFSRCTSVSSLDSPESRSIASVQSEPCSGMVSCIISPSDLPDPSGOTMPPSRSKTTP 1440
Db	
Qy	1441 PPQTQAKREVPKNKAPTAEKRESGPQAAVNAAVQVQVLPDADTLHLHFATESPDGF 1500
Db	
Qy	1441 PPQTQAKREVPKNKAPTAEKRESGPQAAVNAAVQVQVLPDADTLHLHFATESPDGF 1500
Db	
Qy	1501 SCSSSLSALSDELDPFTQKQVELRIMPVQENDNGNETESEQPKESNENKEAEKIDSE 1560
Db	
Qy	1501 SCSSSLSALSDELDPFTQKQVELRIMPVQENDNGNETESEQPKESNENKEAEKIDSE 1560
Db	
Qy	1561 KDLDDDDDDIIELEECIIISAMPTKSSRGKPKAPQATASKLPPPVARKPSQLPVYKLLPS 1620
Db	
Qy	1561 KDLDDDDDDIIELEECIIISAMPTKSSRGKPKAPQATASKLPPPVARKPSQLPVYKLLPS 1620
Db	
Qy	1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGSVRGGAQ 1680
Db	
Qy	1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGSVRGGAQ 1680
Db	
Qy	1681 SGEPEKDDTTPTEGRSTDEAOGKTSSTVTPELDDNKABEGDILAEICINSAMPKASHKP 1740
Db	
Qy	1681 SGEPEKDDTTPTEGRSTDEAOGKTSSTVTPELDDNKABEGDILAEICINSAMPKASHKP 1740
Db	
Qy	1741 FRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYTRVRKNADSKNNLN 1800
Db	
Qy	1741 FRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYTRVRKNADSKNNLN 1800
Db	
Qy	1801 AERVFSNKNKQNLKNKNKDFNDKLPNNEDVRGSAFDSPHHTPIEGTGYCFGRND 1860
Db	
Qy	1801 AERVFSNKNKQNLKNKNKDFNDKLPNNEDVRGSAFDSPHHTPIEGTGYCFGRND 1860
Db	
Qy	1861 SLSSLDLDDDDVLSREKAEKAKENKESEAKVTSHTLTSTNQOSANKTQAIAKQPINR 1920
Db	
Qy	1861 SLSSLDLDDDDVLSREKAEKAKENKESEAKVTSHTLTSTNQOSANKTQAIAKQPINR 1920
Db	
Qy	1921 GQPKFILOKQSTFPQSSKQIDPDRGAATDEKLNQNPATENTVPCFPHNSLSLSLSDIDQNN 1980
Db	
Qy	1921 GQPKFILOKQSTFPQSSKQIDPDRGAATDEKLNQNPATENTVPCFPHNSLSLSLSDIDQNN 1980
Db	
Qy	1981 NKENEPIKETPPPSQGEPSKQVAPKSFHVEDPVCFSRNSLSLSLSLSDIDEDLLQ 2040
Db	
Qy	1981 NKENEPIKETPPPSQGEPSKQVAPKSFHVEDPVCFSRNSLSLSLSLSDIDEDLLQ 2040
Db	

Db 300 PRLLTSHLGTKEVYVYSLLSMLGTHDKDDMGRTLLAMSSQDSCTSMFQSCCLPELLIQLL 359
 QY 361 HGNDKDSVLLGNSRGSKEARARASALHNIIHSQPDCKRGRREIRVLHLLLOIRAYCETC 420
 Db 360 HGNDKDSVLLGNSRGSKEARARASALHNIIHSQPDCKRGRREIRVLHLLLOIRAYCETC 419
 QY 421 NEWQEAHPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEHRHAMNBLGLQIAELLQ 480
 Db 420 NEWQEAHPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEHRHAMNBLGLQIAELLQ 479
 QY 481 VDCENYGTNDHYSITLARYAGMALTNLTFGDVANKATLCMWKCMRALVAQLKSEEDL 540
 Db 480 VDCENYGTNDHYSITLARYAGMALTNLTFGDVANKATLCMWKCMRALVAQLKSEEDL 539
 QY 541 QOVIASVLRLNLSRADVNSKKTLEVGSVKALMECALEVKKESTLKSVLNALWLSAHCT 600
 Db 540 QOVIASVLRLNLSRADVNSKKTLEVGSVKALMECALEVKKESTLKSVLNALWLSAHCT 599
 QY 601 ENKADICAVDQALAFVLGTLTYRSQNTNLTALIESGGGILRVNSSLIATNEDHRQILRENN 660
 Db 600 ENKADICAVDQALAFVLGTLTYRSQNTNLTALIESGGGILRVNSSLIATNEDHRQILRENN 659
 QY 661 CLQTLLOHLKSHSLTVGNACGTLNLSARNPKDOEALWDMGAVSMKQNLHSHKHMAM 720
 Db 660 CLQTLLOHLKSHSLTVGNACGTLNLSARNPKDOEALWDMGAVSMKQNLHSHKHMAM 719
 QY 721 GSAAALRLMANRPKADANIMSPGSSLPVLHVRKQKALAEALDAHLSETFNDNLS 780
 Db 720 GSAAALRLMANRPKADANIMSPGSSLPVLHVRKQKALAEALDAHLSETFNDNLS 779
 QY 781 PKASHRSKORHKQSLYGVDFDTNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSRGSS 840
 Db 780 PKASHRSKORHKQSLYGVDFDTNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSRGSS 839
 QY 841 LDSSSEKDRSLERERGI GLGNYPATENPGTSSKGLQI8TAAQIAKVMEEVSAIHTS 900
 Db 840 LDSSSEKDRSLERERGI GLGNYPATENPGTSSKGLQI8TAAQIAKVMEEVSAIHTS 899
 QY 901 QEDRSSGTTHELCHVTDERNALERSAAHTSNFTYNTFKSENSNPTCSMPYAKLEYKSS 960
 Db 900 QEDRSSGTTHELCHVTDERNALERSAAHTSNFTYNTFKSENSNPTCSMPYAKLEYKSS 959
 QY 961 NDLSNSVSSNDGKRGQMKPSIESYSEDDSKFCYQYQPADLAHKIHSANHMDNDGE 1020
 Db 960 NDLSNSVSSNDGKRGQMKPSIESYSEDDSKFCYQYQPADLAHKIHSANHMDNDGE 1019
 QY 1021 LPTPINYSKLYDEQLNGRQSPQNERWARPKHIEDEIKQSEORQSRNOSTTYPVYTE 1080
 Db 1020 LPTPINYSKLYDEQLNGRQSPQNERWARPKHIEDEIKQSEORQSRNOSTTYPVYTE 1079
 QY 1081 STDDKHLKPOPHFGQECVSPYRSRGANGSETRNVRGSHNGINQVNSQSLCOEDDYEDDKP 1140
 Db 1080 STDDKHLKPOPHFGQECVSPYRSRGANGSETRNVRGSHNGINQVNSQSLCOEDDYEDDKP 1139
 QY 1141 TNYSERYSEEEQHEBERPTNYSIKYNEKRVHDQPIDYSLKYATDIPSSQKQSPFSKGS 1200
 Db 1140 TNYSERYSEEEQHEBERPTNYSIKYNEKRVHDQPIDYSLKYATDIPSSQKQSPFSKGS 1199
 QY 1201 SSGQSKTEHMSSENSTTPSSNAKRONOLHPSSAQSRSGOPOKAAATCKVSSINOETIQ 1260
 Db 1200 SSGQSKTEHMSSENSTTPSSNAKRONOLHPSSAQSRSGOPOKAAATCKVSSINOETIQ 1259
 QY 1261 TYCVEDTPICFRCSLSSLSAEDEIGCNQTTQEADSAANTLQIAEIKGKIGTRSAEDPV 1320
 Db 1260 TYCVEDTPICFRCSLSSLSAEDEIGCNQTTQEADSAANTLQIAEIKGKIGTRSAEDPV 1319
 QY 1321 SEVPAYVQHPRTKSSRLQSSLSSESARHAKAVEFFSGAKSPSKGAOTPKSPPEHYVQET 1380
 Db 1320 SEVPAYVQHPRTKSSRLQSSLSSESARHAKAVEFFSGAKSPSKGAOTPKSPPEHYVQET 1379
 QY 1381 PLMFSRCTSVSSLDSPESFESIASVQSEPCSGMVSGIISPSDLDPSPGQTMPPSRSKTTP 1440

Db 1380 PLMFSRCTSVSSLDSPESFESIASVQSEPCSGMVSGIISPSDLDPSPGQTMPPSRSKTTP 1439
 QY 1441 PPQOTAQTKREVKNKAPTAEKRESQPKQAAVNAAVQVRQVLPDADTLLHFAFESTPDGF 1500
 Db 1440 PPQOTAQTKREVKNKAPTAEKRESQPKQAAVNAAVQVRQVLPDADTLLHFAFESTPDGF 1499
 QY 1501 SCSSSLSALSILDBPFTQKQVELRIMPVQENDNGNETESEQPKESNENQEKAEKTIIDSE 1560
 Db 1500 SCSSSLSALSILDBPFTQKQVELRIMPVQENDNGNETESEQPKESNENQEKAEKTIIDSE 1559
 QY 1561 KDLLDDSDDDDDIIEBECIISAMPTKSSRKKGKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
 Db 1560 KDLLDDSDDDDDIIEBECIISAMPTKSSRKKGKPAQATASKLPPPVARKPSQLPVYKLLPS 1619
 QY 1621 QNLQPKHVSFTPGDDMPRVYCEGTPIINFSTATSLSDLTITBSPNNEIILAAAGVGGGAQ 1680
 Db 1620 QNLQPKHVSFTPGDDMPRVYCEGTPIINFSTATSLSDLTITBSPNNEIILAAAGVGGGAQ 1679
 QY 1681 SGFEKEDTITPTEGRSTDEAOGGKTSVTIPELDDNKABEGDILACINISAMPKGSHKP 1740
 Db 1680 SGFEKEDTITPTEGRSTDEAOGGKTSVTIPELDDNKABEGDILACINISAMPKGSHKP 1739
 QY 1741 FRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYTRVRKNADSKNLN 1800
 Db 1740 FRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYTRVRKNADSKNLN 1799
 QY 1801 AERVFSDNKKSKONLKNRSKDENDKLPNNEDRVGSEAFDPSHHYTPLEGTYCFSRND 1860
 Db 1800 AERVFSDNKKSKONLKNRSKDENDKLPNNEDRVGSEAFDPSHHYTPLEGTYCFSRND 1859
 QY 1861 SLSSLPDDDDDVLSREKAEIRKAKENKSEAVT8HTELTSNQQSANKTQALAKQPINR 1920
 Db 1860 SLSSLPDDDDDVLSREKAEIRKAKENKSEAVT8HTELTSNQQSANKTQALAKQPINR 1919
 QY 1921 GQPKPILOKQSTPQSSKDIIPDRGAATDEKLQNFATENTPVCFSHNSLSLSLSDIOENN 1980
 Db 1920 GQPKPILOKQSTPQSSKDIIPDRGAATDEKLQNFATENTPVCFSHNSLSLSLSDIOENN 1979
 QY 1981 NKNEPIKETEPDSDSGEPKPOASGYAPKSFVEEDTPVCFSRNSLSLSLSDIODELLQ 2040
 Db 1980 NKNEPIKETEPDSDSGEPKPOASGYAPKSFVEEDTPVCFSRNSLSLSLSDIODELLQ 2039
 QY 2041 ECISAMPKPKKPSRLKGDNEKHSRNMWGILOEDTLDKDQRPDSEHGLSPDSENF 2100
 Db 2040 ECISAMPKPKKPSRLKGDNEKHSRNMWGILOEDTLDKDQRPDSEHGLSPDSENF 2099
 QY 2101 WKAIQEGANSIVSLHQAACLSQACSSDSLSLSKSGISLGSPPHLLTPQEEKPFT 2160
 Db 2100 WKAIQEGANSIVSLHQAACLSQACSSDSLSLSKSGISLGSPPHLLTPQEEKPFT 2159
 QY 2161 SNKGPRILKPGKSTLETTKIESESGIKGKKVYKSLITGKVRNSSEISGQMKQPLQAN 2220
 Db 2160 SNKGPRILKPGKSTLETTKIESESGIKGKKVYKSLITGKVRNSSEISGQMKQPLQAN 2219
 QY 2221 MPSISRGRTMIHLPVGNSSSTSPVSKGPPPLKTASKSPSEGQTATTSPRGAKPSVKS 2280
 Db 2220 MPSISRGRTMIHLPVGNSSSTSPVSKGPPPLKTASKSPSEGQTATTSPRGAKPSVKS 2279
 QY 2281 ELSVPARQTSQIOGSSKAPSRSGRSTPSRPAQQLSRPIQSPGRNISIFSGRNGISPPN 2340
 Db 2280 ELSVPARQTSQIOGSSKAPSRSGRSTPSRPAQQLSRPIQSPGRNISIFSGRNGISPPN 2339
 QY 2341 KLSQLPRTSPSTASTKSGSGMKSYTSPORQMSQQNLTKQTLGSKVASSIPRSESASKG 2400
 Db 2340 KLSQLPRTSPSTASTKSGSGMKSYTSPORQMSQQNLTKQTLGSKVASSIPRSESASKG 2399
 QY 2401 LQNMWNGANKKVELSRMSTKSGSESDRSRERPLVROSTTIKEAPSTPLRKLKEESA 2460
 Db 2400 LQNMWNGANKKVELSRMSTKSGSESDRSRERPLVROSTTIKEAPSTPLRKLKEESA 2459
 QY 2461 SFESLSPSSRPASPTRSQATPVLSPLPDMVSLSTHSSVOAGGWRKLPPLNLSPTIEYNDG 2520
 Db 2460 SFESLSPSSRPASPTRSQATPVLSPLPDMVSLSTHSSVOAGGWRKLPPLNLSPTIEYNDG 2519

QY 2521 RPAKHDIARSHSPRLINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSPS 2580
 Db 2520 RPAKHDIARSHSPRLINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSPS 2579
 QY 2581 SEKAKSEDEKXVNSISGTSKQKQVSAKGTWRKIKENEPSPNTSQTVSSGATNGAES 2640
 Db 2580 SEKAKSEDEKXVNSISGTSKQKQVSAKGTWRKIKENEPSPNTSQTVSSGATNGAES 2639
 QY 2641 KTLIYQAPAVSKTDEYVWRIEDCPINPRSGRSPGNTPPVIDSVSEKANPNIKDSKN 2700
 Db 2640 KTLIYQAPAVSKTDEYVWRIEDCPINPRSGRSPGNTPPVIDSVSEKANPNIKDSKN 2699
 QY 2701 QAKQVNGSVPMRTVGLNRLTSFIQVDAPDQKGTGTEIKPGQNNPVPVSETNESPIVERT 2760
 Db 2700 QAKQVNGSVPMRTVGLNRLTSFIQVDAPDQKGTGTEIKPGQNNPVPVSETNESPIVERT 2759
 QY 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPSPRKSSADTSARPQIPTPVNNNTKKRDSKT 2820
 Db 2760 PFSSSSSKHSSPSGTVAAARVTPFNPNPSPRKSSADTSARPQIPTPVNNNTKKRDSKT 2819
 QY 2821 DSTESSGTQPKRHSGSYLVTSV 2843
 Db 2820 DSTESSGTQPKRHSGSYLVTSV 2842

RESULT 15
 ABG90968
 ID ABG90968 standard; peptide; 2842 AA.
 XX AC ABG90968;
 XX DT 29-NOV-2002 (first entry)
 XX DE Human APC protein.
 XX KW Adenomatous polyposis coli; APC; human; neoplastic tissue;
 XX KW mutation detection; tumour; cancer.
 XX OS Homo sapiens.
 XX PN US6413727-B1.
 XX PD 02-JUL-2002.
 XX PF 25-MAY-1995; 95US-00449731.
 XX PR 16-JAN-1991; 91GB-00000962.
 XX PR 16-JAN-1991; 91GB-00000963.
 XX PR 16-JAN-1991; 91GB-00000974.
 XX PR 16-JAN-1991; 91GB-00000975.
 XX PR 08-AUG-1991; 91US-00741940.
 XX PR 12-AUG-1994; 94US-00289548.
 XX (UJGO) UNIV JOHNS HOPKINS.
 XX (UTAH) UNIV UTAH.
 XX (NICA-) JAPANESE FOUND CANCER RES.
 XX (ZENE) ZENECA LTD.
 XX Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ, Joslyn G;
 PI Kinzler K, Markham AF, Nakamura Y, Thliveris A, Vogelstein B;
 PI White RL;
 XX WPI; 2002-641559/69.
 XX Method to aid in the diagnosis/prognosis of neoplastic tissues in humans,
 PT by detecting somatic alteration of wild-type APC protein in tumor tissue
 PT isolated from human, the alteration indicating neoplasia of the tissue.
 XX Example 15; Fig 3; 140pp; English.
 XX This invention relates to a novel method to aid in the diagnosis or
 CC prognosis of a neoplastic tissue of a human. The method involves

CC detecting somatic alteration of wild-type adenomatous polyposis coli)
 CC protein in a tumour tissue isolated from a human (the alteration is
 CC indicating neoplasia of the tissue). The method of the invention is
 CC useful in diagnosis or prognosis of a neoplastic tissue of a human. The
 CC method is useful in detection of genetic predisposition to cancer. The
 CC present sequence represents a peptide sequence used in the method of the
 CC invention
 XX Sequence 2842 AA;
 SQ
 Query Match 99.6%; Score 14515.5; DB 5; Length 2842;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2835; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
 QY 1 MAASDYDOLLKQVAKVENSINLRQELSDNSHLTKLETASNKKEVLKQLQGSIEDEAM 60
 Db 1 MAASDYDOLLKQVAKVENSINLRQELSDNSHLTKLETASNKKEVLKQLQGSIEDEAM 60
 QY 61 ASSGOIDLLERLKEINLDSSNFPVGLKRSKMSLSYSGREGSVSSRSGECPVPMGSPFR 120
 Db 61 ASSGOIDLLERLKEINLDSSNFPVGLKRSKMSLSYSGREGSVSSRSGECPVPMGSPFR 120
 QY 121 RGFVNGSRSTGYLEELKERSILLADLDKEEKQWYVLAQLQNLTKRISLPTENFSL 180
 Db 121 RGFVNGSRSTGYLEELKERSILLADLDKEEKQWYVLAQLQNLTKRISLPTENFSL 179
 QY 181 QTDLTRQLEYEARQIRVAMEEQGLGTQDMKRAQRRIARIQIQIEKDILRIQLQSQAT 240
 Db 180 QTDLTRQLEYEARQIRVAMEEQGLGTQDMKRAQRRIARIQIQIEKDILRIQLQSQAT 239
 QY 241 EAERSSQNKHETGSHDAERQEGQVGEINMATSNGCGSTTRMDHDTASVSSSSHTSA 300
 Db 240 EAERSSQNKHETGSHDAERQEGQVGEINMATSNGCGSTTRMDHDTASVSSSSHTSA 299
 QY 301 PRRLTSHLGTVMVYSLLSMLGTHDXDMSRTLLAMSSSDSCISMRQSGCLPLLIQLL 360
 Db 300 PRRLTSHLGTVMVYSLLSMLGTHDXDMSRTLLAMSSSDSCISMRQSGCLPLLIQLL 359
 QY 361 HGNDKDSVLLGNSRSGKEARASAAALHNIHSOPDDKRRREIRVHLLEQIRAYCETC 420
 Db 360 HGNDKDSVLLGNSRSGKEARASAAALHNIHSOPDDKRRREIRVHLLEQIRAYCETC 419
 QY 421 WEQEAHEPQMDQKPNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLQ 480
 Db 420 WEQEAHEPQMDQKPNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLQ 479
 QY 481 VDCEMYGLTNDHYSITLRYAGMALTNLTGFDVANKATLCMRKGMALVAQKSESEDL 540
 Db 480 VDCEMYGLTNDHYSITLRYAGMALTNLTGFDVANKATLCMRKGMALVAQKSESEDL 539
 QY 541 QOVTASVLRNLWSRADVNSKKTREVGSVKALMECALEVKKESTLKSVLGALNLSAHT 600
 Db 540 QOVTASVLRNLWSRADVNSKKTREVGSVKALMECALEVKKESTLKSVLGALNLSAHT 599
 QY 601 ENKADICAVDQALAFVLGTLYRSQTNTLAIIESGGGILRNVSLLIATNEDHRQILRENN 660
 Db 600 ENKADICAVDQALAFVLGTLYRSQTNTLAIIESGGGILRNVSLLIATNEDHRQILRENN 659
 QY 661 CLQTLLOHLKSHSLTIYSNACGTLWNLSARNPKDQBALWDMGAVSMKLNLSHGHKMIAM 720
 Db 660 CLQTLLOHLKSHSLTIYSNACGTLWNLSARNPKDQBALWDMGAVSMKLNLSHGHKMIAM 719
 QY 721 GSAALRNLMANRPAKYKDANINSPGSSLPFLHVRKQKALEAEADACHLSETFDNIDNLS 780
 Db 720 GSAALRNLMANRPAKYKDANINSPGSSLPFLHVRKQKALEAEADACHLSETFDNIDNLS 779
 QY 781 PKASHRSKQRHKSQLYGDIYVDFTRHDNDNSDPNTGNTVLPYNTTVLPSSSSRSGS 840
 Db 780 PKASHRSKQRHKSQLYGDIYVDFTRHDNDNSDPNTGNTVLPYNTTVLPSSSSRSGS 839
 QY 841 LDSRSKDRSLERERIGLGNVHPATENPCTSKRGLQISTTAAQIAKVMEVSAHTS 900
 Db 840 LDSRSKDRSLERERIGLGNVHPATENPCTSKRGLQISTTAAQIAKVMEVSAHTS 899

Db 1201 SSGOSSKTEHMSSESENSTPSSNAKQKQNLHPSSAQSRGQKQKAAATCKVSIINQETIQ 1260
Qy 1261 TYCVEDTPICFSRCSLSLSLSSAEDEIGCQNTTQEADSANTLQIAIEIKGKIGTRSAEDPV 1320
Db 1261 TYCVEDTPICFSRCSLSLSLSSAEDEIGCQNTTQEADSANTLQIAIEIKGKIGTRSAEDPV 1320
Qy 1321 SEVPANVQHPRTKSSRLQGSLSLSSAEDEIGCQNTTQEADSANTLQIAIEIKGKIGTRSAEDPV 1380
Db 1321 SEVPANVQHPRTKSSRLQGSLSLSSAEDEIGCQNTTQEADSANTLQIAIEIKGKIGTRSAEDPV 1380
Qy 1381 PLMFESRCTSVSSLDSPFSRGIASSVQSEPCGVMVSGIISPSDLPDGPQGTMPSPRSKTPP 1440
Db 1381 PLMFESRCTSVSSLDSPFSRGIASSVQSEPCGVMVSGIISPSDLPDGPQGTMPSPRSKTPP 1440
Qy 1441 PPPQTAQTKREVKNKAPTAKEBESGPKQAANVAORVQVLPDADTLHLHFATESTPDGF 1500
Db 1441 PPPQTAQTKREVKNKAPTAKEBESGPKQAANVAORVQVLPDADTLHLHFATESTPDGF 1500
Qy 1501 SCSSLSALSLSDEFFIOKVELRIMPVQNDNGNETESQPKESNENQEKAEKTIIDSE 1560
Db 1501 SCSSLSALSLSDEFFIOKVELRIMPVQNDNGNETESQPKESNENQEKAEKTIIDSE 1560
Qy 1561 KDLDDSDDDDDIELEBCCIISAMPTKSSRGKKAQPAQASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KDLDDSDDDDDIELEBCCIISAMPTKSSRGKKAQPAQASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOPKHVSFTPGDDMPVYCVGTPINFSTATSLDLTIESPPNELAAGSGVGGQA 1680
Db 1621 QNRLOPKHVSFTPGDDMPVYCVGTPINFSTATSLDLTIESPPNELAAGSGVGGQA 1680
Qy 1681 SGEFEKEDTPTTEGRSTDEAGGKTSSVTIPELDNKAEGDILAEICINSAMPKGSHPK 1740
Db 1681 SGEFEKEDTPTTEGRSTDEAGGKTSSVTIPELDNKAEGDILAEICINSAMPKGSHPK 1740
Qy 1741 FRVKKINDVOQOASASSAPKNOLDGKKKTPSPYKPIPONTEYTRVRKNADSKNNLN 1800
Db 1741 FRVKKINDVOQOASASSAPKNOLDGKKKTPSPYKPIPONTEYTRVRKNADSKNNLN 1800
Qy 1801 AERFVSNKDSKKNLKNNSKDFNDKLPNNEDVRGSAFSDSPHHYTPBEGTYPYCFSRND 1860
Db 1801 AERFVSNKDSKKNLKNNSKDFNDKLPNNEDVRGSAFSDSPHHYTPBEGTYPYCFSRND 1860
Qy 1861 SLSLDPDDDDVLSRKAELRKAENKESAEKVTSHTELTSQOQANKTQATKQPINR 1920
Db 1861 SLSLDPDDDDVLSRKAELRKAENKESAEKVTSHTELTSQOQANKTQATKQPINR 1920
Qy 1921 GQPKPILQKOSTFPQSKDIPDRGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Db 1921 GQPKPILQKOSTFPQSKDIPDRGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Qy 1981 NKNEPIKETEPPDSQGEPSKQASGAPKSFHVETDTPVCFSRNSLSLSLSDIDQENN 2040
Db 1981 NKNEPIKETEPPDSQGEPSKQASGAPKSFHVETDTPVCFSRNSLSLSLSDIDQENN 2040
Qy 2041 ECISAMPKPKKPSRLKDNKESPRNMGILGLEDLTLDKIQRDSEHGLSPDSENF 2100
Db 2041 ECISAMPKPKKPSRLKDNKESPRNMGILGLEDLTLDKIQRDSEHGLSPDSENF 2100
Qy 2101 WKAIQEGANSIVSSLHQAACCLSRQASDSDSILSKSGISLGSGPHLTPQDEKPKFT 2160
Db 2101 WKAIQEGANSIVSSLHQAACCLSRQASDSDSILSKSGISLGSGPHLTPQDEKPKFT 2160
Qy 2161 SNKGPRILKKEKSTLTETKTESKGIKGGKVVYKSLITGKVRNSEISGQMKQPLQAN 2220
Db 2161 SNKGPRILKKEKSTLTETKTESKGIKGGKVVYKSLITGKVRNSEISGQMKQPLQAN 2220
Qy 2221 MPSISRGRTMHIHIGVRNNSSTSPVSKGPPPLKTPASKSPSEGTATTSFPGKXSVKS 2280
Db 2221 MPSISRGRTMHIHIGVRNNSSTSPVSKGPPPLKTPASKSPSEGTATTSFPGKXSVKS 2280
Qy 2281 ELSPVARQTSQIGSSKAPRSGRSDTTPRPAQOPLSRPIQSPGRNSISPGNGISPPN 2340
Db 2281 ELSPVARQTSQIGSSKAPRSGRSDTTPRPAQOPLSRPIQSPGRNSISPGNGISPPN 2340

Qy 2341 KLSQLPRTSSPTASTKSSSGSKMSYTPSGRMSQQNLTKQTGLSKVAASSIPRSESASK 2400
Db 2341 KLSQLPRTSSPTASTKSSSGSKMSYTPSGRMSQQNLTKQTGLSKVAASSIPRSESASK 2400
Qy 2401 LNMQNGNANKKVELSRMSSTKSSGSESDRSEBPVLVROSTFIKEAPSPPTLRKLEESA 2460
Db 2401 LNMQNGNANKKVELSRMSSTKSSGSESDRSEBPVLVROSTFIKEAPSPPTLRKLEESA 2460
Qy 2461 SFESLSPSSRSPASPTRSQAOPTVLSPLPMSLSTHSSVOAGGWRKLPNLSPTIEYNDG 2520
Db 2461 SFESLSPSSRSPASPTRSQAOPTVLSPLPMSLSTHSSVOAGGWRKLPNLSPTIEYNDG 2520
Qy 2521 RPAKHDIARSHSPSPRLPINRSGTWKREHSHSSSLPRVSTWRRTGSSSSILSASSES 2580
Db 2521 RPAKHDIARSHSPSPRLPINRSGTWKREHSHSSSLPRVSTWRRTGSSSSILSASSES 2580
Qy 2581 SEKAKSEDEKRVNISGTQSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKRVNISGTQSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
Qy 2641 KTLIYQAPAVSKTEDVWRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
Db 2641 KTLIYQAPAVSKTEDVWRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
Qy 2701 QAKQNVNGSVPMRTVGLNELTSTFIQVADAPDQXGTBIKPGQNNPVPVSETNESPIVERT 2760
Db 2701 QAKQNVNGSVPMRTVGLNELTSTFIQVADAPDQXGTBIKPGQNNPVPVSETNESPIVERT 2760
Qy 2761 PFSSSSSKHSPSGTVAARVTPFNYPNPSPRKSSADSTARPSPQIPTPVANNTKRDSKT 2820
Db 2761 PFSSSSSKHSPSGTVAARVTPFNYPNPSPRKSSADSTARPSPQIPTPVANNTKRDSKT 2820
Qy 2821 DSTESSGTQSPKSHSGSVLTVTSV 2843
Db 2821 DSTESSGTQSPKSHSGSVLTVTSV 2843

RESULT 11
US-09-003-687A-7
; Sequence 7, Application US/09003687A
; Patent No. 5998600
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/003.687A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,355
; FILING DATE: 20-MAR-1997
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998600e
US-09-003-687A-7

Query Match          99.7%; Score 14533; DB 2; Length 2973;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2936; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAASDYDQLLKQVEALKMENSNLRLQELDNLSNHLTKLETEASNKMKVLLKQLQGSIEDEAM 60
DB 1 MAASDYDQLLKQVEALKMENSNLRLQELDNLSNHLTKLETEASNKMKVLLKQLQGSIEDEAM 60
QY 61 ASSGOIDLLEKLKELNLDSSNPPGVKLRKMSLSYSGREGSVSRSGECSPVPMWGFPR 120
DB 61 ASSGOIDLLEKLKELNLDSSNPPGVKLRKMSLSYSGREGSVSRSGECSPVPMWGFPR 120
QY 121 RGFVNGSRSTGYLFELEKERSLLIADLDKEEKDWTYYAQLNLTKRIDSPLTENFSL 180
DB 121 RGFVNGSRSTGYLFELEKERSLLIADLDKEEKDWTYYAQLNLTKRIDSPLTENFSL 180
QY 181 QTDLTROLEYEAQIRVAMEEQLGTCQDMKRAQRRIARQQIETKOLLRQLQSQAT 240
DB 181 QTDLTROLEYEAQIRVAMEEQLGTCQDMKRAQRRIARQQIETKOLLRQLQSQAT 240
QY 241 EAERSSQNKHETGSHDAERQEGQGVGEINNATSGNGGCGSTRMDHETASVLSSSSTHSA 300
DB 241 EAERSSQNKHETGSHDAERQEGQGVGEINNATSGNGGCGSTRMDHETASVLSSSSTHSA 300
QY 301 PRRLTSHLGTVKVMYVLSLLMGLTHDKDDMGRTLLAMSSODSCISMRQSGCLPLLIQLL 360
DB 301 PRRLTSHLGTVKVMYVLSLLMGLTHDKDDMGRTLLAMSSODSCISMRQSGCLPLLIQLL 360
QY 361 HGNDKDSVLLNGSRGSKARASAAALHNIITHSQDDDKRGREIRVLHLLEQIRAYCETC 420
DB 361 HGNDKDSVLLNGSRGSKARASAAALHNIITHSQDDDKRGREIRVLHLLEQIRAYCETC 420
QY 421 WEWQEAHEFGMDQDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAELLQ 480
DB 421 WEWQEAHEFGMDQDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAELLQ 480
QY 481 VDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALVAQLKSEBDL 540
DB 481 VDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALVAQLKSEBDL 540
QY 541 QQVIALSVLRNLSWRADVNSKTLREVGSKALMECALEVKKESTLKSVLNLSALHACT 600
DB 541 QQVIALSVLRNLSWRADVNSKTLREVGSKALMECALEVKKESTLKSVLNLSALHACT 600
QY 601 ENKADICAVDGLAFVLGTLTVRSQTNTLAIIESGGGILRNWSSLIATNEDHROILRENN 660
DB 601 ENKADICAVDGLAFVLGTLTVRSQTNTLAIIESGGGILRNWSSLIATNEDHROILRENN 660
QY 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARPDKQDALWDVGAVSMKLNLIHSHKXIAM 720
DB 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARPDKQDALWDVGAVSMKLNLIHSHKXIAM 720
QY 721 GSAAALRNLMANRPAYKDANIMSPGSSLPVLHVRKQKALEAELDAQHLSFTFDNIDLS 780
DB 721 GSAAALRNLMANRPAYKDANIMSPGSSLPVLHVRKQKALEAELDAQHLSFTFDNIDLS 780
QY 781 PKASHRSKQRHKQSLGYGVFDNTHDDNRSDNFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840

```

```

781 PKASHRSKQRHKQSLGYGVFDNTHDDNRSDNFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
841 LSSSRSEKORSIERERIGLGNVHPATENPGTSSKRGLOISTTAAQIAKMEEVSAIHTS 900
841 LSSSRSEKORSIERERIGLGNVHPATENPGTSSKRGLOISTTAAQIAKMEEVSAIHTS 900
901 QEDRSSGSGTTELHCVTDERNALRRSSAAATHGNTYNTFKSENSNRCTCSMPYAKLEYKRS 960
901 QEDRSSGSGTTELHCVTDERNALRRSSAAATHGNTYNTFKSENSNRCTCSMPYAKLEYKRS 960
961 NDSLNSVSSNDGYGKQGMKPSIESYSEDDSKFCSYGOYPADLAHIIISANHMDNDGE 1020
961 NDSLNSVSSNDGYGKQGMKPSIESYSEDDSKFCSYGOYPADLAHIIISANHMDNDGE 1020
1021 LDTPIYSLKYSDQLNSGRQSPQNRWARPKHIIEDIKQSEQRQSNQSTTYPVYTE 1080
1021 LDTPIYSLKYSDQLNSGRQSPQNRWARPKHIIEDIKQSEQRQSNQSTTYPVYTE 1080
1081 STDDKHLKTFQPHGQECVSPYRSGANGSETRVNSHGINQNVQSQSLCQEDDYEDDKP 1140
1081 STDDKHLKTFQPHGQECVSPYRSGANGSETRVNSHGINQNVQSQSLCQEDDYEDDKP 1140
1141 TNYSEYSEBEEQHEEBERPTNYSIKYNEKRVHVDQPIDYSLKYATDIPSSQKQSFPSKS 1200
1141 TNYSEYSEBEEQHEEBERPTNYSIKYNEKRVHVDQPIDYSLKYATDIPSSQKQSFPSKS 1200
1201 SSGQSKTEHMSSESTSTPSNNAKQNLHPSSAQSRSGQPQKAATCKVSSINQETIQ 1260
1201 SSGQSKTEHMSSESTSTPSNNAKQNLHPSSAQSRSGQPQKAATCKVSSINQETIQ 1260
1261 TYCVDTPICFSCSSLSLSAEDDEIGNCTOERADSAANTLOIAEIKIGITRSEADPV 1320
1261 TYCVDTPICFSCSSLSLSAEDDEIGNCTOERADSAANTLOIAEIKIGITRSEADPV 1320
1321 SEVPAYSQHPRTKSSRLQSSLSSESARHKAVEFPSSGAKSPKSGAQTPKSPPEHYVQET 1380
1321 SEVPAYSQHPRTKSSRLQSSLSSESARHKAVEFPSSGAKSPKSGAQTPKSPPEHYVQET 1380
1381 PLMFSCSTSVSSLDSPESRSIASSVSEPCSGMWSCIISPSDLDPSPGQTMPPSRKTPP 1440
1381 PLMFSCSTSVSSLDSPESRSIASSVSEPCSGMWSCIISPSDLDPSPGQTMPPSRKTPP 1440
1441 PPQTQATKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPDADTLHFAFATESPDGF 1500
1441 PPQTQATKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPDADTLHFAFATESPDGF 1500
1501 SCSSLSALSLEDPFTQKDVLRIMPVVOENDNGNETESEQPKESNENOEKAEKTI DSE 1560
1501 SCSSLSALSLEDPFTQKDVLRIMPVVOENDNGNETESEQPKESNENOEKAEKTI DSE 1560
1561 KDLLDSDDDDDIILBECIIISAMPTKSSRGKKPAOTASKLPPPVARKPSQLPVYKLLPS 1620
1561 KDLLDSDDDDDIILBECIIISAMPTKSSRGKKPAOTASKLPPPVARKPSQLPVYKLLPS 1620
1621 QNRLOPKHVSFTFGDDMPRVYCVETGPIINFSTATSLDLTIESPPNELAAGEVGRGAQ 1680
1621 QNRLOPKHVSFTFGDDMPRVYCVETGPIINFSTATSLDLTIESPPNELAAGEVGRGAQ 1680
1681 SGFEFKGDTIPTEGRSTDEAOGGKTSSVTIPELDNKAEGDILAEIINSAMPKGSHKP 1740
1681 SGFEFKGDTIPTEGRSTDEAOGGKTSSVTIPELDNKAEGDILAEIINSAMPKGSHKP 1740
1741 FRVKTKMDVQQAASASSAPNKNQLDGKKKPTSPVKPIPQNTYETRVKKNADSKNNLN 1800
1741 FRVKTKMDVQQAASASSAPNKNQLDGKKKPTSPVKPIPQNTYETRVKKNADSKNNLN 1800
1801 AEFVSDNKKQKQNLKNNSKDFNDKLPNNEDVRGSAFDPGPHHTPIEGTPYCFSRND 1860
1801 AEFVSDNKKQKQNLKNNSKDFNDKLPNNEDVRGSAFDPGPHHTPIEGTPYCFSRND 1860
1861 SLSSLDFFDDDDVLSREKAEKRNKSEAKVTSTHETLTSNQQSANKTQALAKOPINR 1920

```

```

Db 1861 SLSSLOFDDDDVLSREKAEELKAKENKSEAKVTSHTLTSNQOSANKTQAIKQPINR 1920
Qy 1921 GQPKPILQKOSTPQSSKOIPDRGAATDEKLNFAIENTVPCFPHNSLSLSIDIDQENN 1980
Db 1921 GQPKPILQKOSTPQSSKOIPDRGAATDEKLNFAIENTVPCFPHNSLSLSIDIDQENN 1980
Qy 1981 NKNEPIKETEPDPSOGEPKSPQASGVAPKSFHVEDTPVCFGRNSLSLSIDSDDLQ 2040
Db 1981 NKNEPIKETEPDPSOGEPKSPQASGVAPKSFHVEDTPVCFGRNSLSLSIDSDDLQ 2040
Qy 2041 ECISAMPKPKKPSRLKGNEXHSPRNMGIIGEDTLTKDQIQRDPSHGLSPSENPD 2100
Db 2041 ECISAMPKPKKPSRLKGNEXHSPRNMGIIGEDTLTKDQIQRDPSHGLSPSENPD 2100
Qy 2101 WKAIQEGANSIVSLHQAANAACLSQAASDSLSLSKSGISLSPFHLTPDQEKPT 2160
Db 2101 WKAIQEGANSIVSLHQAANAACLSQAASDSLSLSKSGISLSPFHLTPDQEKPT 2160
Qy 2161 SNKGPRILKEGKSTLTETKKIESGKIGKGVKYSKSLITGKVRNSSEISGQWKQPLQ 2220
Db 2161 SNKGPRILKEGKSTLTETKKIESGKIGKGVKYSKSLITGKVRNSSEISGQWKQPLQ 2220
Qy 2221 MPISIRGRTHIHPGVNSSSSTSPVSKGPPKLTTPASKSPSEGQTATTSPRGAKPSVK 2280
Db 2221 MPISIRGRTHIHPGVNSSSSTSPVSKGPPKLTTPASKSPSEGQTATTSPRGAKPSVK 2280
Qy 2281 ELSPVARQTSQIGSSKAPSRGSRDSTSPRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
Db 2281 ELSPVARQTSQIGSSKAPSRGSRDSTSPRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSPSTASTKSGSGKMSYTSQGRQMSQNLTKQGLSKXASISPRSSASKG 2400
Db 2341 KLSQLPRTSPSTASTKSGSGKMSYTSQGRQMSQNLTKQGLSKXASISPRSSASKG 2400
Qy 2401 LNOWNNGANKKVELSRMSTSKSGSDSRPVRQSTFIKAPSPPTLRKLEESA 2460
Db 2401 LNOWNNGANKKVELSRMSTSKSGSDSRPVRQSTFIKAPSPPTLRKLEESA 2460
Qy 2461 SPFSLSPSPSPASPTRSQAQTVLPSLPDMSLSTHSSVOAGWRKLPNLSPTIYNDG 2520
Db 2461 SPFSLSPSPSPASPTRSQAQTVLPSLPDMSLSTHSSVOAGWRKLPNLSPTIYNDG 2520
Qy 2521 RPAKRHDIAHSHSPSRLPINSGTWKRSHKSSSLPRVSTWRRTGSSSISLGSSES 2580
Db 2521 RPAKRHDIAHSHSPSRLPINSGTWKRSHKSSSLPRVSTWRRTGSSSISLGSSES 2580
Qy 2581 SEXAKSEDEKXVANSISGTKOSKENOVSAKGTWRKIKENEFSPNTNSTSTVSSGATNGAES 2640
Db 2581 SEXAKSEDEKXVANSISGTKOSKENOVSAKGTWRKIKENEFSPNTNSTSTVSSGATNGAES 2640
Qy 2641 KTLIYQAPAVSKTEVWVRIEDCPINPRSGSPGTNTPPVLDVSEKANPNIKSDKN 2700
Db 2641 KTLIYQAPAVSKTEVWVRIEDCPINPRSGSPGTNTPPVLDVSEKANPNIKSDKN 2700
Qy 2701 QAKQNVGNGVMPRTVGLNRLTSFQVADPDQKTEIKPGQNNPVVSETNESPIVERT 2760
Db 2701 QAKQNVGNGVMPRTVGLNRLTSFQVADPDQKTEIKPGQNNPVVSETNESPIVERT 2760
Qy 2761 PFSSSSSKSSPSGTVAAVTPFNPNPSPRKSSADSTGARPQIPTFNNNTKKEDSKT 2820
Db 2761 PFSSSSSKSSPSGTVAAVTPFNPNPSPRKSSADSTGARPQIPTFNNNTKKEDSKT 2820
Qy 2821 DSTESSGTQSPKXHSGLYTSV 2843
Db 2821 DSTESSGTQSPKXHSGLYTSV 2843

```

RESULT 12
 US-09-136-605-7
 ; Sequence 7, Application us/09136605A
 ; Patent No. 6140052
 ; GENERAL INFORMATION:
 ; APPLICANT: He, Tong-Chuan

```

; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCP-4, and APC Interact to
; FILE REFERENCE: 1107.75741
; CURRENT APPLICATION NUMBER: US/09/136,605A
; EARLIER FILING DATE: 1998-08-20
; EARLIER FILING DATE: 1997-03-20
; EARLIER FILING DATE: 1997-03-20
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2973
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-136-605-7

Query Match 99.7%; Score 14533; DB 3; Length 2973;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAAASYDQLLKQVEALKWNENSLRQELNEDNSHLTKLETEASNKVELKQLOQSIIDEAM 60
Db 1 MAAASYDQLLKQVEALKWNENSLRQELNEDNSHLTKLETEASNKVELKQLOQSIIDEAM 60
Qy 61 ASSQIDILLERLKEINLDSSNFPQVKLRSMKSLRSYSGREGSVSSRGSGCSPVPMGSFPR 120
Db 61 ASSQIDILLERLKEINLDSSNFPQVKLRSMKSLRSYSGREGSVSSRGSGCSPVPMGSFPR 120
Qy 121 RGFVNGRESTGYLEELKERSLLADLDKBEKDWYLAQLQNLTKRIDSLPLTENFSL 180
Db 121 RGFVNGRESTGYLEELKERSLLADLDKBEKDWYLAQLQNLTKRIDSLPLTENFSL 180
Qy 181 QTDLTRQLEVEARQIRVAMEEQLGTQDMKKAQRRIARIQIETKDIRIROLQSOAT 240
Db 181 QTDLTRQLEVEARQIRVAMEEQLGTQDMKKAQRRIARIQIETKDIRIROLQSOAT 240
Qy 241 EAERSSQNKHETGSHDAERONEGQVGEINMATSNGCGSTTRMDHETASVLSSTHSA 300
Db 241 EAERSSQNKHETGSHDAERONEGQVGEINMATSNGCGSTTRMDHETASVLSSTHSA 300
Qy 301 PRRLTSHLGTVMVYSLMLGTHDXXDMSRTLLAMSSSDSCISMRSQGCPLLIQLL 360
Db 301 PRRLTSHLGTVMVYSLMLGTHDXXDMSRTLLAMSSSDSCISMRSQGCPLLIQLL 360
Qy 361 HGNDKDSVLLGNSRGSKEARASAAALHNTIHSQDDPKGRREIRVLHLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNSRGSKEARASAAALHNTIHSQDDPKGRREIRVLHLEQIRAYCETC 420
Qy 421 WEMQEAHEPGDQDKNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAIELLQ 480
Db 421 WEMQEAHEPGDQDKNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAIELLQ 480
Qy 481 VDCEMYGLTDHYHYSITLRRYAGMALTNLTGFDVANKATLCSMKGCMALVAQLKSEEDL 540
Db 481 VDCEMYGLTDHYHYSITLRRYAGMALTNLTGFDVANKATLCSMKGCMALVAQLKSEEDL 540
Qy 541 QQVIASVLRNLSWADVNSKKTLEBVSVKALMECALEVKESTLKSVLNLSNLSAHC 600
Db 541 QQVIASVLRNLSWADVNSKKTLEBVSVKALMECALEVKESTLKSVLNLSNLSAHC 600
Qy 601 ENKADICAVDGAFLVGTTLTYRSQTNTLAIIESGGGILNVSLSLIATNEDHRLRENN 660
Db 601 ENKADICAVDGAFLVGTTLTYRSQTNTLAIIESGGGILNVSLSLIATNEDHRLRENN 660
Qy 661 CLQTLQHLKSHSLTIIVSNACGTLWNLSARNPKDQEBALWDMGAVSMKXNLSKXMIAM 720
Db 661 CLQTLQHLKSHSLTIIVSNACGTLWNLSARNPKDQEBALWDMGAVSMKXNLSKXMIAM 720
Qy 721 GSAALRNLMANRPAPKVDANIMSPGSSLPVSRKQKALEAEIDAQHLSETFDNIDNLS 780

```


721 GSAALRLNLRNPAKYKADANINSGSLPSLHVRRKQKALEAEALDAQHLGETFTDNIDNLS 780
781 PKASHRSKORHQSLYGDYVFTNRHDDNRDNFTGNMTVLSPYANTVLPSSSSSRG 840
781 PKASHRSKORHQSLYGDYVFTNRHDDNRDNFTGNMTVLSPYANTVLPSSSSSRG 840
841 LDSRSEKORSLERIGLGNYPATENPOTSSKRGLOISTTAAQIAKWEVSAIHTS 900
841 LDSRSEKORSLERIGLGNYPATENPOTSSKRGLOISTTAAQIAKWEVSAIHTS 900
901 QEDRSSGTTLHCVTDERNALRSSAAHCHSNFTNFKSENRTCSMPYAKLEYKRSS 960
901 QEDRSSGTTLHCVTDERNALRSSAAHCHSNFTNFKSENRTCSMPYAKLEYKRSS 960
961 NDSLNSVSSNDGYKRGQMKPSIESYSEDDSKFCYQGPADLAHKIHSANHMDNDGE 1020
961 NDSLNSVSSNDGYKRGQMKPSIESYSEDDSKFCYQGPADLAHKIHSANHMDNDGE 1020
1021 LDTINTYSLKSDQLNSGROSQONERWAPKHIIEDEIKOSEQROSRNQSTIYPYTE 1080
1021 LDTINTYSLKSDQLNSGROSQONERWAPKHIIEDEIKOSEQROSRNQSTIYPYTE 1080
1081 STDJXHLKFQHFQOQECVSPYRSGANGSETNVRGSHNGINQVOSLQEDDDYEDDKP 1140
1081 STDJXHLKFQHFQOQECVSPYRSGANGSETNVRGSHNGINQVOSLQEDDDYEDDKP 1140
1141 TNYSERYSERQHEERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSFSEKS 1200
1141 TNYSERYSERQHEERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSFSEKS 1200
1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAOSRGOPOKAATCKVSSINQETIQ 1260
1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAOSRGOPOKAATCKVSSINQETIQ 1260
1261 TYCVEDTPICFSRCSLSLSAEDIGCNQTTQEADSNANTLOIAEIKIGITSAEDPV 1320
1261 TYCVEDTPICFSRCSLSLSAEDIGCNQTTQEADSNANTLOIAEIKIGITSAEDPV 1320
1321 SEVFAVSOHPTKSRLOGSSLSSESARHKAVEPPSGAKSKSGAOTPKSPPHYVOET 1380
1321 SEVFAVSOHPTKSRLOGSSLSSESARHKAVEPPSGAKSKSGAOTPKSPPHYVOET 1380
1381 PLMFSRCTSVSLSDFSRSSTASSVQSEPCSGMWGIIISPSDLDPSPQOTMPPSRKSTPP 1440
1381 PLMFSRCTSVSLSDFSRSSTASSVQSEPCSGMWGIIISPSDLDPSPQOTMPPSRKSTPP 1440
1441 PPTQACTKREVPKNAKPTAEKRESGPKQAANVAQVQVLPDADTLILHFAESTPDGF 1500
1441 PPTQACTKREVPKNAKPTAEKRESGPKQAANVAQVQVLPDADTLILHFAESTPDGF 1500
1501 SCSSLSALSLEDEPIQDVELRIMPVQENDNGNETESEOPKESNENQKEAEKTTDSE 1560
1501 SCSSLSALSLEDEPIQDVELRIMPVQENDNGNETESEOPKESNENQKEAEKTTDSE 1560
1561 KDLDDSDDDIEILEECIIISAMPTKSRKKGKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
1561 KDLDDSDDDIEILEECIIISAMPTKSRKKGKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
1621 QNLQOPKHVSFTPCDDMPRYVCVEGTPIINSTATSLDILTIESPPNELAAGEVGRGAQ 1680
1621 QNLQOPKHVSFTPCDDMPRYVCVEGTPIINSTATSLDILTIESPPNELAAGEVGRGAQ 1680
1681 SGEFEKRTIITEGRSTDEAOGKTSVVTIPELDDNKAEEGDIIAECINSAMPKGSHPK 1740
1681 SGEFEKRTIITEGRSTDEAOGKTSVVTIPELDDNKAEEGDIIAECINSAMPKGSHPK 1740
1741 FRVKIMQVOQASASSAPKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKNLN 1800
1741 FRVKIMQVOQASASSAPKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKNLN 1800
1801 AERFVSDNKKQNLKNNKDFNDKLPNNEDRVGRGSFAFDSPPHYTPIEGTFCFGRND 1860
1801 AERFVSDNKKQNLKNNKDFNDKLPNNEDRVGRGSFAFDSPPHYTPIEGTFCFGRND 1860

RESULT 13

US-07-741-940-7

; Sequence 7, Application US/07741940

1861 SLSSLDFFDDDDVLDGREKAELKAKENKESAKVTSHTELTNQQSANKTOAIKOPINR 1920
1861 SLSSLDFFDDDDVLDGREKAELKAKENKESAKVTSHTELTNQQSANKTOAIKOPINR 1920
1921 GPKPILQKOSTPPOSSKDIIDRGAATDBKLOFAIENTPVCFSHNSLSLSLSDIDQENN 1980
1921 GPKPILQKOSTPPOSSKDIIDRGAATDBKLOFAIENTPVCFSHNSLSLSLSDIDQENN 1980
1981 NKNEPIKETEPPDSDGSEPSKQAGYAPKSPHVEDTPVCFSRNSSLSSLSLSDIDQ 2040
1981 NKNEPIKETEPPDSDGSEPSKQAGYAPKSPHVEDTPVCFSRNSSLSSLSLSDIDQ 2040
2041 ECITSSAMPKKKXPSRLKGDNEKHSPPNMGIIIGEDLTLDLKDIQRPDSHSGSPSENFD 2100
2041 ECITSSAMPKKKXPSRLKGDNEKHSPPNMGIIIGEDLTLDLKDIQRPDSHSGSPSENFD 2100
2101 WKAIQEGANSIVSSLHQAAAAACLSQASSSDSDSILSLKSGISLSPHLPDQEKPT 2160
2101 WKAIQEGANSIVSSLHQAAAAACLSQASSSDSDSILSLKSGISLSPHLPDQEKPT 2160
2161 SNKGPRILKPGKSTLTETKKIESEKGIKGGKVKYKSLITGKVRNSSEISGQMKPQLOAN 2220
2161 SNKGPRILKPGKSTLTETKKIESEKGIKGGKVKYKSLITGKVRNSSEISGQMKPQLOAN 2220
2221 MPSISGRMTMIIHPGVNRSSSTSPVSKGPKLTPKSPSEGGTATTPRGAKPSVK 2280
2221 MPSISGRMTMIIHPGVNRSSSTSPVSKGPKLTPKSPSEGGTATTPRGAKPSVK 2280
2281 ELSPVARQTSQIGGSSKAPSRGSRDSTPSRPAQPLSRPIOSPGRNSISPGRNGISPN 2340
2281 ELSPVARQTSQIGGSSKAPSRGSRDSTPSRPAQPLSRPIOSPGRNSISPGRNGISPN 2340
2341 KLSQLPRTSPSTASTKSSGSGKMSYTSRQMSQONLTQOTGLSKNASSIPRSESASK 2400
2341 KLSQLPRTSPSTASTKSSGSGKMSYTSRQMSQONLTQOTGLSKNASSIPRSESASK 2400
2401 LQNMNNGANKVLESRMSTKSSGSESDRBERPVLVQSTFIKEAPSTLRRKLEESA 2460
2401 LQNMNNGANKVLESRMSTKSSGSESDRBERPVLVQSTFIKEAPSTLRRKLEESA 2460
2461 SFESLSPSRSPASPTRSQACTPVLSPSLPMSLSHSSVQAGGWRKLPNLSPTEYNDG 2520
2461 SFESLSPSRSPASPTRSQACTPVLSPSLPMSLSHSSVQAGGWRKLPNLSPTEYNDG 2520
2521 RPAKRHDIAHSHSESRLPINRSGTWKREHSHSSLPVSTWRTTSGSSSILSASSES 2580
2521 RPAKRHDIAHSHSESRLPINRSGTWKREHSHSSLPVSTWRTTSGSSSILSASSES 2580
2581 SEKAKSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSPNTNSTSVSSGATNGAES 2640
2581 SEKAKSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSPNTNSTSVSSGATNGAES 2640
2641 KTLIIYQVAPAVSKTEVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
2641 KTLIIYQVAPAVSKTEVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
2701 QAKQNGVSGVWRTVGLNRLTSPFOVAPQKGTIETKPGQNNPVVSETNESPIVERT 2760
2701 QAKQNGVSGVWRTVGLNRLTSPFOVAPQKGTIETKPGQNNPVVSETNESPIVERT 2760
2761 PFSSSSSKHSGSPGTVAAARVTFNTNPNPRKSSADSTGARSQIPTPVNNNTKKEDSKT 2820
2761 PFSSSSSKHSGSPGTVAAARVTFNTNPNPRKSSADSTGARSQIPTPVNNNTKKEDSKT 2820
2821 DSTESSGTSQPKHSGSYLVTSV 2843
2821 DSTESSGTSQPKHSGSYLVTSV 2843

; Patent No. 5352775
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILIP J.
 ; APPLICANT: JOSLYN, GEOFF
 ; APPLICANT: KINZLER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: NAKAMURA, YUSUKE
 ; APPLICANT: THLIVERIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 94
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner, Birch, McKie & Beckett
 ; STREET: 1001 G Street, NW
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001-4598
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/741,940
 ; FILING DATE: 19920109
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kagan, Sarah A.
 ; REGISTRATION NUMBER: 32,141
 ; REFERENCE/DOCKET NUMBER: 1107.035574
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-508-9100
 ; TELEFAX: 202-508-9289
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2842 amino acids
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: APC
 ;

Query Match 99.6%; Score 14515.5; DB 1; Length 2842;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2835; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
 QY 1 MAAASYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKMKVQLQGSIEDEAM 60
 DB 1 MAAASYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKMKVQLQGSIEDEAM 60
 QY 61 ASSGQIDLLERLKLNDSSNFFGVKLRKMSLRSGSREGSVSRGSCSVPMGSPFR 120
 DB 61 ASSGQIDLLERLKLNDSSNFFGVKLRKMSLRSGSREGSVSRGSCSVPMGSPFR 120
 QY 121 RGVNFSRESTGYLELEKERSLLDLDKEEKEKDWYIAQLNLTKRIDSPLTENFSL 180
 DB 121 RGVNFSRESTGYLELEKERSLLDLDKEEKEKDWYIAQLNLTKRIDSPLTENFSL 179
 QY 181 QTLTRQLREYARQIRVAMEEQLGTCDQMEKRAQRRIARIQIEKDIIRQLQSQAT 240
 DB 180 QTMTRQLREYARQIRVAMEEQLGTCDQMEKRAQRRIARIQIEKDIIRQLQSQAT 239
 QY 241 EAERSQNHETGSHDAERQNEQGVGEINMATSNGQGGSTTRMDHETASVLSSSSTHSA 300

Db 240 EAERSQNHETGSHDAERQNEQGVGEINMATSNGQGGSTTRMDHETASVLSSSSTHSA 299
 QY 301 PRLTSHLGTKEVMYVLLSMLGTHDKDDSRITLAMSSSQDSCISMRSQSGCLPLLIQLL 360
 Db 300 PRLTSHLGTKEVMYVLLSMLGTHDKDDSRITLAMSSSQDSCISMRSQSGCLPLLIQLL 359
 QY 361 HGNDKDSVLLGNSRSGKEARASAAALHNIHSDPDKRGRREIRVLHLLLEQIRAYCETC 420
 Db 360 HGNDKDSVLLGNSRSGKEARASAAALHNIHSDPDKRGRREIRVLHLLLEQIRAYCETC 419
 QY 421 WEQEAHEPGMDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 480
 Db 420 WEQEAHEPGMDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 479
 QY 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTGFDVANKATLCMKGCMRALVAQLKSESEDL 540
 Db 480 VDCEMYGLTNDHYSITLRRYAGMALTNLTGFDVANKATLCMKGCMRALVAQLKSESEDL 539
 QY 541 QOVIASVRLNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSIVLSALNLSAHC 600
 Db 540 QOVIASVRLNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSIVLSALNLSAHC 599
 QY 601 ENKADICAVDGAFLVGTTLTYRSQTNLTALIESGGILRNVSLLIATNEDHRQILRENN 660
 Db 600 ENKADICAVDGAFLVGTTLTYRSQTNLTALIESGGILRNVSLLIATNEDHRQILRENN 659
 QY 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKDQEBALWDGAVSMKNIHSHKHMIAM 720
 Db 660 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKDQEBALWDGAVSMKNIHSHKHMIAM 719
 QY 721 GSAALNLSWANRPKADANIMSPGSSLSPLHVRKQKALEAELDAQHLETDNDNLNLS 780
 Db 720 GSAALNLSWANRPKADANIMSPGSSLSPLHVRKQKALEAELDAQHLETDNDNLNLS 779
 QY 781 PKASHRSKQRHKQSLGYDYVDFNRRDDNRSDFNTGNMTVLSPYLNTTVPSSSSSRGS 840
 Db 780 PKASHRSKQRHKQSLGYDYVDFNRRDDNRSDFNTGNMTVLSPYLNTTVPSSSSSRGS 839
 QY 841 LDSSRSKDRSLERERIGIGNYHPATENBCTSKRGLQISTTAAQIAKMEVESAHTS 900
 Db 840 LDSSRSKDRSLERERIGIGNYHPATENBCTSKRGLQISTTAAQIAKMEVESAHTS 899
 QY 901 QEDRSSGSTTTLHCVTDERNALRRSSAAHTSHNTYFTKSENENRITCMTPIAKLEYKRSS 960
 Db 900 QEDRSSGSTTTLHCVTDERNALRRSSAAHTSHNTYFTKSENENRITCMTPIAKLEYKRSS 959
 QY 961 NDSLNSVSSNDGYKRGQMKPSIESYSEDDDESFCYQYPADLAHKIHSANHMDNDGE 1020
 Db 960 NDSLNSVSSNDGYKRGQMKPSIESYSEDDDESFCYQYPADLAHKIHSANHMDNDGE 1019
 QY 1021 LDTPIYSLKYSDQLNSGRQSPQNERWARPKHIDEIKQSEQRQSRNQSTTYPVYTE 1080
 Db 1020 LDTPIYSLKYSDQLNSGRQSPQNERWARPKHIDEIKQSEQRQSRNQSTTYPVYTE 1079
 QY 1081 STDDKHLKFPHFQGCQVCVPSYRSGANGSETNRVGNHGINQVNSQSLQCEDDYEDDKP 1140
 Db 1080 STDDKHLKFPHFQGCQVCVPSYRSGANGSETNRVGNHGINQVNSQSLQCEDDYEDDKP 1139
 QY 1141 TNYSERVSEBQHEEBERPTNYSIKYNEEKRHVDOPIDYSLKYATDIPSSQKOSFQFSKS 1200
 Db 1140 TNYSERVSEBQHEEBERPTNYSIKYNEEKRHVDOPIDYSLKYATDIPSSQKOSFQFSKS 1199
 QY 1201 SSGQSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260
 Db 1200 SSGQSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1259
 QY 1261 TYCVEDTPICFRSCSSLSLSAEDDEIGCNOTTOEADSAANTLOIAETKIGITRSEDV 1320
 Db 1260 TYCVEDTPICFRSCSSLSLSAEDDEIGCNOTTOEADSAANTLOIAETKIGITRSEDV 1319
 QY 1321 SEVPAVSOHPRTKSSRLQSSLSSESARHKAVERFPSSGAKSPSKGAOTPKSPPEHYVQET 1380
 Db 1320 SEVPAVSOHPRTKSSRLQSSLSSESARHKAVERFPSSGAKSPSKGAOTPKSPPEHYVQET 1379

QY 1381 PLMFGRCTSVSSLDSEFERSIASSVQSEPCSGMWGIIISPSDLPSDPCQTWPPRSKTPP 1440
DB 1380 PLMFGRCTSVSSLDSEFERSIASSVQSEPCSGMWGIIISPSDLPSDPCQTWPPRSKTPP 1439
QY 1441 PPQTAQTKREVPKNAKPTAEKRESGPKQAAVNAARVQVVLDPADATLLHLPATSTPDGF 1500
DB 1440 PPQTAQTKREVPKNAKPTAEKRESGPKQAAVNAARVQVVLDPADATLLHLPATSTPDGF 1499
QY 1501 SCSSLSALSLEDEPIQKDVLRIMPPVQENDNGNETESEOPKESNENOEKAEKTIIDSE 1560
DB 1500 SCSSLSALSLEDEPIQKDVLRIMPPVQENDNGNETESEOPKESNENOEKAEKTIIDSE 1559
QY 1561 KDLDDSDDDDIIEECIISAMPTKSRGKKPAQTASKLPPPVVARKPSOLPVYKLLPS 1620
DB 1560 KDLDDSDDDDIIEECIISAMPTKSRGKKPAQTASKLPPPVVARKPSOLPVYKLLPS 1619
QY 1621 QNRLOPQKHVSTPDDDDPRVYCVGEGTPIINSTATSLDLTIESPPNELAEGEVGRGAQ 1680
DB 1620 QNRLOPQKHVSTPDDDDPRVYCVGEGTPIINSTATSLDLTIESPPNELAEGEVGRGAQ 1679
QY 1681 SGEFEKRDITPEGRSTDEAOGGKTSVVTIPELDDNKAEEGDIILAEICINSAMPKGSHP 1740
DB 1680 SGEFEKRDITPEGRSTDEAOGGKTSVVTIPELDDNKAEEGDIILAEICINSAMPKGSHP 1739
QY 1741 FRVKIMDOVQOASASSAPKNQKLDGKKKTSVPKPIQNTYRTVRKNADSKNNLN 1800
DB 1740 FRVKIMDOVQOASASSAPKNQKLDGKKKTSVPKPIQNTYRTVRKNADSKNNLN 1799
QY 1801 AERFVSDNKKQKQKQKNDKPNNDNRVGRGSPAFSPHHYTIETGTPYCFSRND 1860
DB 1800 AERFVSDNKKQKQKQKNDKPNNDNRVGRGSPAFSPHHYTIETGTPYCFSRND 1859
QY 1861 SLSSLDFFDDDDVLDREKAEKAKENKESAKVTSHTELTSNQQOSANKTOAIKQPNR 1920
DB 1860 SLSSLDFFDDDDVLDREKAEKAKENKESAKVTSHTELTSNQQOSANKTOAIKQPNR 1919
QY 1921 GQPKPILQKQSTPPOSSKDIIPRGAADEKLNQFAIENTPVCFSHNSLSLSDIDQENN 1980
DB 1920 GQPKPILQKQSTPPOSSKDIIPRGAADEKLNQFAIENTPVCFSHNSLSLSDIDQENN 1979
QY 1981 NKENEPIKETEPDQSGEPSPKQASGYAPKSPHVEDTTPVCFSRNSLSLSIDSEDDLLQ 2040
DB 1980 NKENEPIKETEPDQSGEPSPKQASGYAPKSPHVEDTTPVCFSRNSLSLSIDSEDDLLQ 2039
QY 2041 ECISAMPKKKXPSRLKGNEXHSPRNMGGIILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
DB 2040 ECISAMPKKKXPSRLKGNEXHSPRNMGGIILGEDLTLDLKDIOQPDSEHGLSPDSENF 2099
QY 2101 WKAIQEGANSIVSSLHQAAAAACLRQASSDSDSILSLKSGISLSPHLLTPDQEKPEFT 2160
DB 2100 WKAIQEGANSIVSSLHQAAAAACLRQASSDSDSILSLKSGISLSPHLLTPDQEKPEFT 2159
QY 2161 SNKGPRILKPGKSTLETTKIESGKIGKGVKYSKITGKVRNSNISEIGQMQLQAN 2220
DB 2160 SNKGPRILKPGKSTLETTKIESGKIGKGVKYSKITGKVRNSNISEIGQMQLQAN 2219
QY 2221 MFSISGRMTIHTPGVRNSSSTSPVSKGPKLPTPASKSPSEGOTATTPRGAKPSVKS 2280
DB 2220 MFSISGRMTIHTPGVRNSSSTSPVSKGPKLPTPASKSPSEGOTATTPRGAKPSVKS 2279
QY 2281 ELSPVARQTSQIGGSKAPSRGSRDSTPSRPAQPLSRPIQSPGRNISIPGRNGISPPN 2340
DB 2280 ELSPVARQTSQIGGSKAPSRGSRDSTPSRPAQPLSRPIQSPGRNISIPGRNGISPPN 2339
QY 2341 KLSOLPRTSPSTASTKSSGSGMWSTSPGRWQOONLTQGLSKNASSIPRSESASK 2400
DB 2340 KLSOLPRTSPSTASTKSSGSGMWSTSPGRWQOONLTQGLSKNASSIPRSESASK 2399
QY 2401 LNMOMNNGANKKVELSRMSSTKSSGESDRSERPVLVQSTFIKEAPSTLRRKLEESA 2460
DB 2400 LNMOMNNGANKKVELSRMSSTKSSGESDRSERPVLVQSTFIKEAPSTLRRKLEESA 2459

QY 2461 SPESLSPSRPASPTRSQAOQTPVLSPLPDMSLSTHSSVQAGWRKLPNLSPTIEYNDG 2520
DB 2460 SPESLSPSRPASPTRSQAOQTPVLSPLPDMSLSTHSSVQAGWRKLPNLSPTIEYNDG 2519
QY 2521 RPAKRHDIAASHSESRLPINESCTWKREHSKSSSLPRVSTWRRTGSSSSILSASSES 2580
DB 2520 RPAKRHDIAASHSESRLPINESCTWKREHSKSSSLPRVSTWRRTGSSSSILSASSES 2579
QY 2581 SEKAKSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
DB 2580 SEKAKSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2639
QY 2641 KTLIYQAPAVSKTDDVWRIEDCPINNPRSGSPGNTPTPVIDSVSEKANPNIKOSKDN 2700
DB 2640 KTLIYQAPAVSKTDDVWRIEDCPINNPRSGSPGNTPTPVIDSVSEKANPNIKOSKDN 2699
QY 2701 QAKQNVGNGVPMRTVGLNRLTSFIQVDAPDQKGTIIPGQNNPVVSESTNESPIVERT 2760
DB 2700 QAKQNVGNGVPMRTVGLNRLTSFIQVDAPDQKGTIIPGQNNPVVSESTNESPIVERT 2759
QY 2761 PFSSSSSSKHSFSGTVAARVTPFNPNPRKSSADSTGARPQIPTPVNNNTKKRDSKT 2820
DB 2760 PFSSSSSSKHSFSGTVAARVTPFNPNPRKSSADSTGARPQIPTPVNNNTKKRDSKT 2819
QY 2821 DSTESSGTQSPKHSGLVLTSV 2843
DB 2820 DSTESSGTQSPKHSGLVLTSV 2842

RESULT 14

US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids

1981 NKENPIKETEPDPSQGPSPKQAGYAPKPFHVEDTPVCFSRNSSLSLSDSEDDLLQ 2040
1980 NKENPIKETEPDPSQGPSPKQAGYAPKPFHVEDTPVCFSRNSSLSLSDSEDDLLQ 2039
2041 ECISAMPKKKPKSLKGNKHSRNNGGILGDLTLDDKDIORPDEHGLSPDSENF 2100
2040 ECISAMPKKKPKSLKGNKHSRNNGGILGDLTLDDKDIORPDEHGLSPDSENF 2099
2101 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSLSLKSGISLSPHLPDQEKPF 2160
2100 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSLSLKSGISLSPHLPDQEKPF 2159
2161 SNKGPRILKPKGKSTLETKKIESEKGIKGGKVKYKSLITGKVRNSNISGOMKQLOAN 2220
2160 SNKGPRILKPKGKSTLETKKIESEKGIKGGKVKYKSLITGKVRNSNISGOMKQLOAN 2219
2221 MFSISGRRTMIHIPGVRNSSSTSPVSKKGPPLKTPASKSPSEGGTATTSRGAKPVS 2280
2220 MFSISGRRTMIHIPGVRNSSSTSPVSKKGPPLKTPASKSPSEGGTATTSRGAKPVS 2279
2281 ELSPVARQTSQIGGSKAPRSRSDSTPSRPAQOPLSRPTQSPGRNISIFGRNGISPPN 2340
2280 ELSPVARQTSQIGGSKAPRSRSDSTPSRPAQOPLSRPTQSPGRNISIFGRNGISPPN 2339
2341 KLSQLPRTSSPTASTKSGSGKMSYTPSGROMSQNLTKOTGLSKNASSIPRSESASG 2400
2340 KLSQLPRTSSPTASTKSGSGKMSYTPSGROMSQNLTKOTGLSKNASSIPRSESASG 2399
2401 LNMNNGANKKVELSRMSSTKSGSDSRERPVLRQSTFIKEAPSPILRRKLEBSA 2460
2400 LNMNNGANKKVELSRMSSTKSGSDSRERPVLRQSTFIKEAPSPILRRKLEBSA 2459
2461 SPESLSPSRPASPTRSOAQTVPVLSPLPDMSLTHSSVQAGWKLPPNLSPIEYNDG 2520
2460 SPESLSPSRPASPTRSOAQTVPVLSPLPDMSLTHSSVQAGWKLPPNLSPIEYNDG 2519
2521 RPAKRHDIAHRSSEPSRLPINRSGTWKREKSHSSLPVSTWRTGSSSIILSASSES 2580
2520 RPAKRHDIAHRSSEPSRLPINRSGTWKREKSHSSLPVSTWRTGSSSIILSASSES 2579
2581 SEKASEDEKHVNSTSGTKQKENVSAKGTWTKIKENEFSTNSTSTVSGATNGAES 2640
2580 SEKASEDEKHVNSTSGTKQKENVSAKGTWTKIKENEFSTNSTSTVSGATNGAES 2639
2641 KTLIQMAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKSDKN 2700
2640 KTLIQMAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKSDKN 2699
2701 QAKQVNGSVPMRTVGLNRLTSTFIQVDAPQKGTETKPGQNNPVVSETNESPIVERT 2760
2700 QAKQVNGSVPMRTVGLNRLTSTFIQVDAPQKGTETKPGQNNPVVSETNESPIVERT 2759
2761 PFSSSSSKHSSPSTGTVAAARVTPNNVPSRKSADSTSRPSQIPTPVNNNTKKRDSKT 2820
2760 PFSSSSSKHSSPSTGTVAAARVTPNNVPSRKSADSTSRPSQIPTPVNNNTKKRDSKT 2819
2821 DSTESGTSQPKRHSGSYLVTSV 2843
2820 DSTESGTSQPKRHSGSYLVTSV 2842

RESULT 15
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 99.6%; Score 14515.5; DB 1; Length 2842;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2835; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MAAASYDQLLKQVEALKQVENSRLRQELDNLSNHLTKLETSANNKEVILKQSGSIEDEAM 60
Db 1 MAAASYDQLLKQVEALKQVENSRLRQELDNLSNHLTKLETSANNKEVILKQSGSIEDEAM 60
Qy 61 ASSQCIDLLERLKLNDSSNFPQVLRKRWLSRYSYSGREGSVSSRSGECPVPMGSFPR 120
Db 61 ASSQCIDLLERLKLNDSSNFPQVLRKRWLSRYSYSGREGSVSSRSGECPVPMGSFPR 120
Qy 121 RGFVNGSRESTGYLEELKERSLLADLDKEKEKDWYQAQLQNLTKRIDSLPTENFSL 180
Db 121 RGFVNGSRESTGYLEELKERSLLADLDKEKEKDWYQAQLQNLTKRIDSL-LTENFSL 179
Qy 181 QTDLTRQLEVEARQIRVAMEEQGLCTODMEKRAQRIARIQIQIEKDLIRIQLQSQAT 240
Db 180 QTDLTRQLEVEARQIRVAMEEQGLCTODMEKRAQRIARIQIQIEKDLIRIQLQSQAT 239
Qy 241 EAERSSQNKHETGSHDAERQNEGQVGEINNVATSGNQGQSTTRMDHETASVLSSTHSA 300
Db 240 EAERSSQNKHETGSHDAERQNEGQVGEINNVATSGNQGQSTTRMDHETASVLSSTHSA 299
Qy 301 PRRLTSHLTKVENVYLLSMLGTHDKDDMSRILLAMSSQSDSCISMRQSGCPLLIQLL 360
Db 300 PRRLTSHLTKVENVYLLSMLGTHDKDDMSRILLAMSSQSDSCISMRQSGCPLLIQLL 359

QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHSOPDDKRGREIRVTLHLLEQIRAYCETC 420
 Db 360 HGNDKDSVLLGNSRGSKEARASAAALHNIHSOPDDKRGREIRVTLHLLEQIRAYCETC 419
 QY 421 WEMQEAHEPGMDQDNMPAPVEHQICPACVCLMKLSFDEHRHANNELGGLQAI AELLQ 480
 Db 420 WEMQEAHEPGMDQDNMPAPVEHQICPACVCLMKLSFDEHRHANNELGGLQAI AELLQ 479
 QY 481 VDCWYGLTNDHYSITLRRVAGMALTNLTGPDVANKATLCSMKGCMRALVAQIKSSEDL 540
 Db 480 VDCWYGLTNDHYSITLRRVAGMALTNLTGPDVANKATLCSMKGCMRALVAQIKSSEDL 539
 QY 541 QOVLASVLRNLWRADVNSKKTLEVGSVKALMECALEVKESTLKSVALNLSAHLCT 600
 Db 540 QOVLASVLRNLWRADVNSKKTLEVGSVKALMECALEVKESTLKSVALNLSAHLCT 599
 QY 601 ENKADICAVDICALAFIVGTITYRSQNTWLAIIESGGGILRNVSLLIATNEDHRIQILRENN 660
 Db 600 ENKADICAVDICALAFIVGTITYRSQNTWLAIIESGGGILRNVSLLIATNEDHRIQILRENN 659
 QY 661 CLOTLLQHLKSHSLTIVSNACGLTNLNSARNPKDOBALWDMGAVSMLKNLIHSHKHMIAM 720
 Db 660 CLOTLLQHLKSHSLTIVSNACGLTNLNSARNPKDOBALWDMGAVSMLKNLIHSHKHMIAM 719
 QY 721 GSAAALNLANRPAKYKANDINIMSGSLPSLHVRCOKALEAEALDAQHLSETFDNIDNLS 780
 Db 720 GSAAALNLANRPAKYKANDINIMSGSLPSLHVRCOKALEAEALDAQHLSETFDNIDNLS 779
 QY 781 PKASHRSKORHKQSLYGDYVDTNRHDDNDSFNCTGNVTLPYNTNTVLPSSSSRGS 840
 Db 780 PKASHRSKORHKQSLYGDYVDTNRHDDNDSFNCTGNVTLPYNTNTVLPSSSSRGS 839
 QY 841 LDDSRSEKDRSLERERIGLGNHPATENPGTSSKRGLOISTTAAQIAKYMEEVSAIHTS 900
 Db 840 LDDSRSEKDRSLERERIGLGNHPATENPGTSSKRGLOISTTAAQIAKYMEEVSAIHTS 899
 QY 901 QEDRSSSTTELHCVTDERNALRRSSAAHSTHNTYNTFTKSENRRCPMPYAKLEYKRS 960
 Db 900 QEDRSSSTTELHCVTDERNALRRSSAAHSTHNTYNTFTKSENRRCPMPYAKLEYKRS 959
 QY 961 NDSLNSVSSNDGYGKRGOMKPSIESYSEDDSKFCVGYQPADLAHKIHSANHMDDNDGE 1020
 Db 960 NDSLNSVSSNDGYGKRGOMKPSIESYSEDDSKFCVGYQPADLAHKIHSANHMDDNDGE 1019
 QY 1021 LDTPIYNSLYKSDOLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSNQSTTYPVYTE 1080
 Db 1020 LDTPIYNSLYKSDOLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSNQSTTYPVYTE 1079
 QY 1081 STDDKHLKFPQFQOQCVSPYRSRANGSETNRVGSNHCINONVCSLCOEDDYEDDKP 1140
 Db 1080 STDDKHLKFPQFQOQCVSPYRSRANGSETNRVGSNHCINONVCSLCOEDDYEDDKP 1139
 QY 1141 TNSYERYSEBEQHEEERPTNYGIKNEEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1200
 Db 1140 TNSYERYSEBEQHEEERPTNYGIKNEEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1199
 QY 1201 SSGOSSXTEHWSSESSENTSPSSNAKQNLHPSSAQSRGOGOKAATCKVASINQETIQ 1260
 Db 1200 SSGOSSXTEHWSSESSENTSPSSNAKQNLHPSSAQSRGOGOKAATCKVASINQETIQ 1259
 QY 1261 TYCVEDTPICFRCSLSSLSAEDICGNQTTQEADSNANTLQIABIKIGITRISAEDPV 1320
 Db 1260 TYCVEDTPICFRCSLSSLSAEDICGNQTTQEADSNANTLQIABIKIGITRISAEDPV 1319
 QY 1321 SEYPAVNSOHRPTKSSRLQCSLSLSESNARKHAFVFPAGKSPSKSGAOTPKSPHYVQET 1380
 Db 1320 SEYPAVNSOHRPTKSSRLQCSLSLSESNARKHAFVFPAGKSPSKSGAOTPKSPHYVQET 1379
 QY 1381 PLMFSTRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIIPSDLPDPSGQTMPPSRSKTPP 1440
 Db 1380 PLMFSTRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIIPSDLPDPSGQTMPPSRSKTPP 1439
 QY 1441 PFPQTAQTKREVFPKNKAPTAKESGPKQAANAAVQVQLPDAITLLHFATESTPDGF 1500

Db 1440 PFPQTAQTKREVFPKNKAPTAKESGPKQAANAAVQVQLPDAITLLHFATESTPDGF 1499
 QY 1501 SCSSLSALSALDEPFIQKDVLRIMPPVQENDNGNETESEBQPKESNENQEKAEKTI DSE 1560
 Db 1500 SCSSLSALSALDEPFIQKDVLRIMPPVQENDNGNETESEBQPKESNENQEKAEKTI DSE 1559
 QY 1561 KOLLDDSDDDDIIELEECIISAMPTKSRKCKKPAQATASKLPPPVARKPSOLPYKLLPS 1620
 Db 1560 KOLLDDSDDDDIIELEECIISAMPTKSRKCKKPAQATASKLPPPVARKPSOLPYKLLPS 1619
 QY 1621 QNRLOPQKHVSFTPGDDMPRVYCVGEGTPIINFSTATSLSDLTIESPPNELAAGEGVGGAQ 1680
 Db 1620 QNRLOPQKHVSFTPGDDMPRVYCVGEGTPIINFSTATSLSDLTIESPPNELAAGEGVGGAQ 1679
 QY 1681 SGEPFKRDTIIPTEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEACINSAMPKGSHPK 1740
 Db 1680 SGEPFKRDTIIPTEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEACINSAMPKGSHPK 1739
 QY 1741 FRVKKIMDOVQOASASSAPNKNOLDGKKKXPTSPVXPIPONTYRTRVRKNADSKNLN 1800
 Db 1740 FRVKKIMDOVQOASASSAPNKNOLDGKKKXPTSPVXPIPONTYRTRVRKNADSKNLN 1799
 QY 1801 AERVSFNDKSKQNLKNNSKDFNDKLPNNEDRVGSAFDSPHHYTPIEGTPYCFSRND 1860
 Db 1800 AERVSFNDKSKQNLKNNSKDFNDKLPNNEDRVGSAFDSPHHYTPIEGTPYCFSRND 1859
 QY 1861 SLSSLDPDDDVDILSRKAEALRKAKENKESAKVTSHLTSTNOCSANKTOQAIKQPINR 1920
 Db 1860 SLSSLDPDDDVDILSRKAEALRKAKENKESAKVTSHLTSTNOCSANKTOQAIKQPINR 1919
 QY 1921 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLQNFPAIENTPVCFSHNSLSLSLSDIDQENN 1980
 Db 1920 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLQNFPAIENTPVCFSHNSLSLSLSDIDQENN 1979
 QY 1981 NKNEPIKETEPFPPDSQGEPSKQASGVAPKSFHYEDTVCFSRNSLSLSLSDIDQENN 2040
 Db 1980 NKNEPIKETEPFPPDSQGEPSKQASGVAPKSFHYEDTVCFSRNSLSLSLSDIDQENN 2039
 QY 2041 ECISAMPEKKKSRLLKGDNEKHSRNGGILGSDLTLDLKDIOQPDSEHGLSPDSSENF 2100
 Db 2040 ECISAMPEKKKSRLLKGDNEKHSRNGGILGSDLTLDLKDIOQPDSEHGLSPDSSENF 2099
 QY 2101 WKATOEGANGSTVSSLHOAAAAAALSRQASDSDSILSKGILSGSPFHLTPDOEEKPFT 2160
 Db 2100 WKATOEGANGSTVSSLHOAAAAAALSRQASDSDSILSKGILSGSPFHLTPDOEEKPFT 2159
 QY 2161 SNKGPRILKPKQEKSTLTKTIESKGIKGGKVKYKSLITGKVRNSSEISGQMKQIPQAN 2220
 Db 2160 SNKGPRILKPKQEKSTLTKTIESKGIKGGKVKYKSLITGKVRNSSEISGQMKQIPQAN 2219
 QY 2221 MPSISRGRTMTHIPGVNNSSSSTSPVSKGPPPLKTPASKSPSEGOATTPSPRGAKPSVKS 2280
 Db 2220 MPSISRGRTMTHIPGVNNSSSSTSPVSKGPPPLKTPASKSPSEGOATTPSPRGAKPSVKS 2279
 QY 2281 ELSVPAQRTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Db 2280 ELSVPAQRTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPN 2339
 QY 2341 KLSQLPRTSSPSTASTKSSGSKMSYTPSGQMSQNLTKQTLGSKNASSIPRSESASKG 2400
 Db 2340 KLSQLPRTSSPSTASTKSSGSKMSYTPSGQMSQNLTKQTLGSKNASSIPRSESASKG 2399
 QY 2401 LNQWNGANGANKVELSRMSTKSSGSESDRSEPVLRQSTFIKEAPSTFLRKKEESA 2460
 Db 2400 LNQWNGANGANKVELSRMSTKSSGSESDRSEPVLRQSTFIKEAPSTFLRKKEESA 2459
 QY 2461 SFESLSPSSRSPASPTRQAQTPVLSPSLPMSLSTSHSVQAGGWRKLPPLNLSPTIENDG 2520
 Db 2460 SFESLSPSSRSPASPTRQAQTPVLSPSLPMSLSTSHSVQAGGWRKLPPLNLSPTIENDG 2519
 QY 2521 RPAKHDIARSHSSPRLPNTNRSGTWKREHSHSSSLPVSSTWRTPGSSSSTILSASSES 2580

Db	2520	RPKRHDIARSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES	2579
Qy	2581	SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES	2640
Db	2580	SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES	2639
Qy	2641	KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKDN	2700
Db	2640	KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKDN	2699
Qy	2701	QAKQNVGNGSVPMETVGLNRLTSFIOVDAPDOKGTEIKQONNPVPVSETNESPIVERT	2760
Db	2700	QAKQNVGNGSVPMETVGLNRLTSFIOVDAPDOKGTEIKQONNPVPVSETNESPIVERT	2759
Qy	2761	PFSSSSSKHSSPSGTVAAARVTPFNYPNPSRKSSADSTARSQIPTPVNNNTKKRDSKT	2820
Db	2760	PFSSSSSKHSSPSGTVAAARVTPFNYPNPSRKSSADSTARSQIPTPVNNNTKKRDSKT	2819
Qy	2821	DSTESSGTQSPKRHSGSYLVTSV	2843
Db	2820	DSTESSGTQSPKRHSGSYLVTSV	2842

Search completed: August 25, 2004, 17:25:42
 Job time : 53.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:06:26 ; Search time 59 Seconds

(without alignments)
4635.127 Million cell updates/sec

Title: US-09-442-489F-7

Perfect score: 14566

Sequence: 1 MAASVDQLLKQVEALKMEN.....ESSGTQSPKRGSSYLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR.78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14566	100.0	2843	1 RBHUAP	adenomatous polyposis coli protein - human
2	13178	90.5	2845	2 T30336	adenomatous polyposis coli protein - human
3	3512.5	24.1	2274	2 T30258	adenomatous polyposis coli protein - human
4	1780.5	12.2	2415	2 T13825	adenomatous polyposis coli protein - human
5	649	4.5	5327	2 T13584	adenomatous polyposis coli protein - human
6	517	3.5	2271	2 P30073	adenomatous polyposis coli protein - human
7	513	3.5	2232	2 T34434	adenomatous polyposis coli protein - human
8	511.5	3.5	1186	2 T23327	adenomatous polyposis coli protein - human
9	509.5	3.3	1188	2 T23330	adenomatous polyposis coli protein - human
10	476	3.3	2774	2 A43359	adenomatous polyposis coli protein - human
11	454	3.1	3507	2 T34513	adenomatous polyposis coli protein - human
12	453	3.1	4776	2 E95206	adenomatous polyposis coli protein - human
13	426.5	2.9	1791	2 T02345	adenomatous polyposis coli protein - human
14	420	2.9	2938	2 T30249	adenomatous polyposis coli protein - human
15	417.5	2.9	2722	2 T20532	adenomatous polyposis coli protein - human
16	415.5	2.9	2738	2 E88320	adenomatous polyposis coli protein - human
17	409	2.8	3942	2 T42730	adenomatous polyposis coli protein - human
18	405.5	2.8	3488	2 T34418	adenomatous polyposis coli protein - human
19	405	2.8	2187	2 T30826	adenomatous polyposis coli protein - human
20	404	2.8	2688	2 I49477	adenomatous polyposis coli protein - human
21	400	2.7	2526	2 T20531	adenomatous polyposis coli protein - human
22	399	2.7	3924	2 S37431	adenomatous polyposis coli protein - human
23	395.5	2.7	3187	2 T20537	adenomatous polyposis coli protein - human
24	395	2.7	5170	2 T15348	adenomatous polyposis coli protein - human
25	394.5	2.7	3147	2 T18674	adenomatous polyposis coli protein - human
26	394	2.7	6713	2 B99921	adenomatous polyposis coli protein - human
27	393.5	2.7	5105	2 T32650	adenomatous polyposis coli protein - human
28	391.5	2.7	3562	2 A47171	adenomatous polyposis coli protein - human
29	379.5	2.6	2248	2 A35938	adenomatous polyposis coli protein - human

30	378.5	2.6	2253	2 T30336	nuclear/mitotic ap
31	377	2.6	1459	2 T32271	hypothetical prote
32	376.5	2.6	2176	2 T1806	toucan gene protei
33	376	2.6	971	2 T19431	hypothetical prote
34	375.5	2.6	2364	2 A56577	microtubule-associ
35	374.5	2.6	4377	2 A55575	ankyrin 3, long sp
36	374	2.6	3256	2 A48666	cell proliferation
37	372	2.6	2649	2 T51023	hypothetical prote
38	371.5	2.6	1560	2 T02885	peroxisome prolife
39	370.5	2.5	2346	2 T18829	nuclear receptor c
40	370	2.5	2453	2 S60254	Basoon protein -
41	369	2.5	3938	2 T42761	breast cancer tumo
42	367	2.5	3328	2 T30835	trithorax homolog
43	367	2.5	3968	2 A44265	hypothetical prote
44	365	2.5	2218	2 B84683	hypothetical prote
45	364.5	2.5	990	2 I51618	nucleolar phosphop

ALIGNMENTS

RESULT 1

RBHUAP
adenomatous polyposis coli protein - human
N;Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1991 #sequence revision 31-Dec-1991 #text_change 21-Jul-2000
C;Accession: A37261; B39658; A49319; I54271
R;Kinzler, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith, C.; Kinzler, K.W.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.
Science 253, 661-665, 1991
A;Title: Identification of FAP locus genes from chromosome 5q21.
A;Reference number: A37261; MUID:91335210; PMID:1651562
A;Accession: A37261
A;Molecule type: mRNA
A;Residues: 1-2843 <KIN>
A;Cross-references: GB:M74088; NID:9182396; PIDN:AAA03586.1; PID:9182397
R;Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grode, arrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Leif Cell 66, 601-613, 1991
A;Title: Identification of deletion mutations and three new genes at the familial polyposis A;Reference number: A39658; MUID:91330307; PMID:1678319
A;Accession: B39658
A;Molecule type: DNA
A;Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P', A;Cross-references: GB:M73548; NID:9190163; PIDN:AAA60354.1; PID:9190164
R;Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelstein Cancer Res. 52, 643-645, 1992
A;Title: Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a c A;Reference number: A44928; MUID:92119623; PMID:1310068
A;Accession: A44928
A;Molecule type: DNA
A;Residues: 1506-1525 <MIK>
A;Cross-references: GB:S78214; NID:9243541; PIDN:AAB21145.1; PID:9243542
A;Note: sequence extracted from NCBI backbone (NCBI:78214, NCBIP:78218)
R;Spirio, L.; Olshwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelbert Cell 75, 951-957, 1993
A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.
A;Reference number: A49319; MUID:94073973; PMID:8252630
A;Accession: A49319
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'G', 143-171, 'P', 173-179 <SPI>
A;Cross-references: GB:S67787; NID:9461061; PIDN:AAD13997.1; PID:94261697
R;Lambertz, S.; Ballhausen, W.G.
Hum. Genet. 90, 650-652, 1993
A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly A;Reference number: I54271; MUID:93186137; PMID:8383094
A;Accession: I54271
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4 <LAW>
A;Cross-references: GB:S56365; NID:9266243; PIDN:AAD14918.1; PID:94262770

C;Genetics:		
A;Gene:	GDB:APC	
A;Cross-references:	GDB:119682; OMIM:175100	
A;Map position:	5q21-5q22	
A;Note:	mutations of this gene can result in familial adenomatous polyposis or sporadic	
C;Superfamily:	adenomatous polyposis coli protein	
C;Keywords:	cancer; familial adenomatous polyposis; tumor suppressor	
F;1-730/Domain:	leucine-rich <NTD>	
F;7-72/Region:	coil #status predicted	
F;185-227/Region:	coil #status predicted	
F;731-2832/Domain:	serine-rich <CTD>	
F;1131-1156/Region:	acidic	
F;1558-1577/Region:	acidic	
F;1866-1893/Region:	highly charged	
Query Match		100.0%; Score 14566; DB 1; Length 2843;
Best Local Similarity		100.0%; Pred. No. 0;
Matches 2843; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	MAAASYDOLLKQVEALKWENSRLPOLEDNSNHLTKYLEASNMKEVLKOLQSGTDEAM 60
Db	1	MAAASYDOLLKQVEALKWENSRLPOLEDNSNHLTKYLEASNMKEVLKOLQSGTDEAM 60
Qy	61	ASSGQIDLLERLKLNDSSNFFGVKLRKMSLRYSYSGREGSVSRGSCSPVPMGSPFR 120
Db	61	ASSGQIDLLERLKLNDSSNFFGVKLRKMSLRYSYSGREGSVSRGSCSPVPMGSPFR 120
Qy	121	RGFNGSRESTGYLEBEKERSLLADLKEEKEKDYVYAOQLNLTKRIDSPLTENPSL 180
Db	121	RGFNGSRESTGYLEBEKERSLLADLKEEKEKDYVYAOQLNLTKRIDSPLTENPSL 180
Qy	181	QDTMTROLEYEARQIRVAMEEOLGTCDQMEKRAQRIARIOIEKDILIRIQLQSOAT 240
Db	181	QDTMTROLEYEARQIRVAMEEOLGTCDQMEKRAQRIARIOIEKDILIRIQLQSOAT 240
Qy	241	EABRSQNHETGSHDAERQEGQVGEINMATSGNGQSTTRMDHETASVLSSTHSA 300
Db	241	EABRSQNHETGSHDAERQEGQVGEINMATSGNGQSTTRMDHETASVLSSTHSA 300
Qy	301	PRLTHSLGTPKPMVYSLSLMGLTHDKDMSRTLLAWSSQSCISMRQSGCLPILLIQL 360
Db	301	PRLTHSLGTPKPMVYSLSLMGLTHDKDMSRTLLAWSSQSCISMRQSGCLPILLIQL 360
Qy	361	HGNDKDSVLGNRSGKEARASAAHNIHSPQDDKRGREIRVHLLEQIRAYCETC 420
Db	361	HGNDKDSVLGNRSGKEARASAAHNIHSPQDDKRGREIRVHLLEQIRAYCETC 420
Qy	421	WEOEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQAELLIQ 480
Db	421	WEOEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQAELLIQ 480
Qy	481	VCCEMYGLTNDHYISITLRYAGMALTNLTFDGVANKATLCMKGWRALVAQLKSESD 540
Db	481	VCCEMYGLTNDHYISITLRYAGMALTNLTFDGVANKATLCMKGWRALVAQLKSESD 540
Qy	541	QOVIASVLRNLWRADVNSKTLREVGSVKALMECALEVKKESTLKSVALWNLSAHT 600
Db	541	QOVIASVLRNLWRADVNSKTLREVGSVKALMECALEVKKESTLKSVALWNLSAHT 600
Qy	601	ENKADICAVDGAFLVGLTYRSQNTLAIIESGGILRNYSLSIATNEDHROLRENN 660
Db	601	ENKADICAVDGAFLVGLTYRSQNTLAIIESGGILRNYSLSIATNEDHROLRENN 660
Qy	661	CLQTLQHLKSHSLTIVSNACGTLWNLSARNPKDQBALWDMGAVSLKMLIHSKHKMIAM 720
Db	661	CLQTLQHLKSHSLTIVSNACGTLWNLSARNPKDQBALWDMGAVSLKMLIHSKHKMIAM 720
Qy	721	GSAALRNLMANEPKAKDANTMSPGSSLPVLRVKKOKALELDAQHLSETFDNIDLS 780
Db	721	GSAALRNLMANEPKAKDANTMSPGSSLPVLRVKKOKALELDAQHLSETFDNIDLS 780
Qy	781	PKASHRSKQRHKOSLYGDIYVDFVTRNDRNDNPNFTGNMTVLSPLYLNTVLPSSSSSRGS 840
Db	781	PKASHRSKQRHKOSLYGDIYVDFVTRNDRNDNPNFTGNMTVLSPLYLNTVLPSSSSSRGS 840
Qy	841	LDSSRSKDRSLERERIGLGNHYHPATENPCTSSKRGQLQISTTAAQIAKWEVSAIHTS 900
Db	841	LDSSRSKDRSLERERIGLGNHYHPATENPCTSSKRGQLQISTTAAQIAKWEVSAIHTS 900
Qy	901	QEDRSSGTTTLCVTDERNALRRSSAAHNSHTNTYNTFKSENSNRTCSMPYAKLEYKRSS 960
Db	901	QEDRSSGTTTLCVTDERNALRRSSAAHNSHTNTYNTFKSENSNRTCSMPYAKLEYKRSS 960
Qy	961	NDLSNVSSSDYGKRGOMKPSIESYSEDDDESKFCSYGOYPADLAHKIHSANHMDNDGE 1020
Db	961	NDLSNVSSSDYGKRGOMKPSIESYSEDDDESKFCSYGOYPADLAHKIHSANHMDNDGE 1020
Qy	1021	LDTPINYSKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Db	1021	LDTPINYSKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Qy	1081	STDDKHLKFOHFGQOECVSPYRSRGANGSETNRVGNHGINQVNSOSLCEQDDYEDDKP 1140
Db	1081	STDDKHLKFOHFGQOECVSPYRSRGANGSETNRVGNHGINQVNSOSLCEQDDYEDDKP 1140
Qy	1141	TNYSERYSEEEHHEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKOSFSPKS 1200
Db	1141	TNYSERYSEEEHHEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKOSFSPKS 1200
Qy	1201	SSGOSSTEHMSSSENSTPSSNAKRONOLHPSSAQSRSQPOKATCKVSSINQETIQ 1260
Db	1201	SSGOSSTEHMSSSENSTPSSNAKRONOLHPSSAQSRSQPOKATCKVSSINQETIQ 1260
Qy	1261	TYCVEDTPICFSRCSLSLSAEDDEIGCQNTQEADSANTLOIAETKEIGTRSAEDPV 1320
Db	1261	TYCVEDTPICFSRCSLSLSAEDDEIGCQNTQEADSANTLOIAETKEIGTRSAEDPV 1320
Qy	1321	SEVPAVQHPRTKSRRLQSSLSSESARHKAIVFSSGAKSPKSGAQTPKSPPEHYVQET 1380
Db	1321	SEVPAVQHPRTKSRRLQSSLSSESARHKAIVFSSGAKSPKSGAQTPKSPPEHYVQET 1380
Qy	1381	PLMFSRCTSVSSLDSPFSRSIASSVQSEPCSGMVSGIISPSDLPDGPQTMPSRSKTPP 1440
Db	1381	PLMFSRCTSVSSLDSPFSRSIASSVQSEPCSGMVSGIISPSDLPDGPQTMPSRSKTPP 1440
Qy	1441	PPPTAQTKREVKNKAPTAEKESGPKQAAVNAAVQVQVLPDADTLHLHFAESTPDGF 1500
Db	1441	PPPTAQTKREVKNKAPTAEKESGPKQAAVNAAVQVQVLPDADTLHLHFAESTPDGF 1500
Qy	1501	SCSSLSALSALDEPFIQKDVLEIRIWPVQNDNGNETESEQPKESNENQKEAEKTIIDSE 1560
Db	1501	SCSSLSALSALDEPFIQKDVLEIRIWPVQNDNGNETESEQPKESNENQKEAEKTIIDSE 1560
Qy	1561	KDLDDDDDDDIIELEECIIISAMPTKSRKAKKPAQTAASKLPPPVARKPSQLPVYKLLPS 1620
Db	1561	KDLDDDDDDDIIELEECIIISAMPTKSRKAKKPAQTAASKLPPPVARKPSQLPVYKLLPS 1620
Qy	1621	QNRLOPKHYVFPDGDMPRVYCVGCTPINFSTATSLSDLTIESPNEIAAGVGGAQ 1680
Db	1621	QNRLOPKHYVFPDGDMPRVYCVGCTPINFSTATSLSDLTIESPNEIAAGVGGAQ 1680
Qy	1681	SGEPEKEDTIPTEGRSTDEAQQGKTSSVTIPELDDNKAESGDIILAEICINSAMPKGSHPK 1740
Db	1681	SGEPEKEDTIPTEGRSTDEAQQGKTSSVTIPELDDNKAESGDIILAEICINSAMPKGSHPK 1740
Qy	1741	FRVKIMQVQOQASASSAENKQLOGKKKPTSPVKPIQNTYTRVTRVKNADSKNIN 1800
Db	1741	FRVKIMQVQOQASASSAENKQLOGKKKPTSPVKPIQNTYTRVTRVKNADSKNIN 1800
Qy	1801	AERVFSNKKQKQNLKNSKDFNDKLPNNEDVRGSPAFDSPHVTPTEGTYCFSRND 1860
Db	1801	AERVFSNKKQKQNLKNSKDFNDKLPNNEDVRGSPAFDSPHVTPTEGTYCFSRND 1860
Qy	1861	SLSLDPLDDDDVLSREKAEIRKAKENKSEAKVTSHTELTSTNQOQANKTQAIKQPINR 1920
Db	1861	SLSLDPLDDDDVLSREKAEIRKAKENKSEAKVTSHTELTSTNQOQANKTQAIKQPINR 1920

1921 QGPKPILOKQSTFPQSSKIDPDRGAATDEKLQNPATENTVCFSHNSLSLSDIDQENN 1980
 Db 1921 QGPKPILOKQSTFPQSSKIDPDRGAATDEKLQNPATENTVCFSHNSLSLSDIDQENN 1980
 Qy 1981 NKENPEIKETEPDPSQCEPKPQASGVAPKSFHVEDTPVCFSRNSSLSDISDEDDLQ 2040
 Db 1981 NKENPEIKETEPDPSQCEPKPQASGVAPKSFHVEDTPVCFSRNSSLSDISDEDDLQ 2040
 Qy 2041 ECISAMPKPKKPKRLKGDNEKHSFPMNGGTLGDLTLDLKDQRPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKPKKPKRLKGDNEKHSFPMNGGTLGDLTLDLKDQRPDSEHGLSPDSENF 2100
 Qy 2101 WKAIQEGANSTVSLHQAACLSROASDSDSLKSGISLGSFPHTPQOEKPF 2160
 Db 2101 WKAIQEGANSTVSLHQAACLSROASDSDSLKSGISLGSFPHTPQOEKPF 2160
 Qy 2161 SNKGPRILKPEKSTLTETKTESKGIKGGKVKYKSLITGKVRNSEISGQMKQPLQAN 2220
 Db 2161 SNKGPRILKPEKSTLTETKTESKGIKGGKVKYKSLITGKVRNSEISGQMKQPLQAN 2220
 Qy 2221 MPSISRGRTMIHIFGVANSSTSPVSKKGPPLKTPASKSPSEGTATTSPRGAKPSVK 2280
 Db 2221 MPSISRGRTMIHIFGVANSSTSPVSKKGPPLKTPASKSPSEGTATTSPRGAKPSVK 2280
 Qy 2281 ELSVARTSOTIGSSKAPSGSRDSTPPRPAOQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Db 2281 ELSVARTSOTIGSSKAPSGSRDSTPPRPAOQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Qy 2341 KLSQLPRTSSPSTASTKSSGSKNYSYTPGRQMSQQLTKQTGLSKNASSIPRSESASK 2400
 Db 2341 KLSQLPRTSSPSTASTKSSGSKNYSYTPGRQMSQQLTKQTGLSKNASSIPRSESASK 2400
 Qy 2401 LQNMNGANKVELSRMSSTKSGESDRSERPVLVROSTTFKEAPSPTLRRKLEESA 2460
 Db 2401 LQNMNGANKVELSRMSSTKSGESDRSERPVLVROSTTFKEAPSPTLRRKLEESA 2460
 Qy 2461 SFESLSPSSRRPASPTRSQAQTPVLSPLPDMSLSTHSSVQAGWRKLPPLNLSPTIEYNDG 2520
 Db 2461 SFESLSPSSRRPASPTRSQAQTPVLSPLPDMSLSTHSSVQAGWRKLPPLNLSPTIEYNDG 2520
 Qy 2521 RPAKHDIARSHSPSRPLPNRSGTWKREHSHSSSLPRVSTWRRTGSSSSILSASES 2580
 Db 2521 RPAKHDIARSHSPSRPLPNRSGTWKREHSHSSSLPRVSTWRRTGSSSSILSASES 2580
 Qy 2581 SEKAKEDEKVNSTSGTKQSKENQVSAKGTWRKI KENERFPTNSTQTQVSSGATNGAES 2640
 Db 2581 SEKAKEDEKVNSTSGTKQSKENQVSAKGTWRKI KENERFPTNSTQTQVSSGATNGAES 2640
 Qy 2641 KTLIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKDN 2700
 Db 2641 KTLIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKDN 2700
 Qy 2701 QAKQVNGSVPMRTVGLNLSFIQVDPQKGTETKPKQNNPVPVSETNESIVERT 2760
 Db 2701 QAKQVNGSVPMRTVGLNLSFIQVDPQKGTETKPKQNNPVPVSETNESIVERT 2760
 Qy 2761 PFSSSSSKHSPSGTVAARTVFPNPNPSPKSADSTSAEPSQITPPVNNNTKRDST 2820
 Db 2761 PFSSSSSKHSPSGTVAARTVFPNPNPSPKSADSTSAEPSQITPPVNNNTKRDST 2820
 Qy 2821 DSTESSGTQSPKRHSGSVLTVSV 2843
 Db 2821 DSTESSGTQSPKRHSGSVLTVSV 2843

R;Su, L.
 Science 256, 668-670, 1992
 A;Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the
 A;Reference number: I49505; MUID:92263101; PMID:1350108
 A;Accession: I49505
 A;Status: preliminary; translated from GB/EMBL/DBDJ
 A;Molecule type: mRNA
 A;Residues: 1-2845 <RES>
 A;Cross-references: GB:M88127; NID:gl91991; PIDN:AAB59632.1; PID:gl91992
 C;Superfamily: adenomatous polyposis coli protein

Query Match 90.5%; Score 13178; DB 2; Length 2845;
 Best Local Similarity 90.2%; Pred. No. 0;
 Matches 2572; Conservative 111; Mismatches 155; Indels 12; Gaps 10;

Qy 1 MAASVYDQLLKQVEALKMENSNLROELEDNSNHLTKLETEASNMEVULKQLOGSIDEAM 60
 Db 1 MAASVYDQLLKQVEALKMENSNLROELEDNSNHLTKLETEASNMEVULKQLOGSIDEAM 60
 Qy 61 ASSGOIDLLERLKEINLNDSSNFPVKLRKMSLRYSVSGREGSVSRSGECSPVPMGSPFR 120
 Db 61 -TSGQIDLLERLKEINLNDSSNFPVKLRKMSLRYSVSGREGSVSRSGECSPVPMGSPFR 118
 Qy 121 RGFVNGSRESTGYLBELEKERSLLADLDKEKEKDWYVAQQLNLTKRIDSPLTENFSL 180
 Db 119 RGFVNGSRESTGYLBELEKERSLLADLDKEKEKDWYVAQQLNLTKRIDSPLTENFSL 178
 Qy 181 QDTMTROLEYEARQIRVAMEEQLGTCQDMEXEAQRIARIQOIEKDIIRIQLLOSQAT 240
 Db 179 QDTMTROLEYEARQIRVAMEEQLGTCQDMEXEAQRIARIQOIEKDIIRIQLLOSQAA 238
 Qy 241 EAERSQNKHETGSHDAERQNEGGVGEINMATSGNGQSTTMDHETASVLSSESSTHA 300
 Db 239 EAERSQNRHDAASHEAGQHEGHGVAESNTAASSGQSPATRVDHETASVLSSESSTHA 298
 Qy 301 PRRLTSHLGTQVEMVYVSLLSMLGTHDKDMRTLLAMSSQSDSCISMROSGCLPLLIQLL 360
 Db 299 PRRLTSHLGTQVEMVYVSLLSMLGTHDKDMRTLLAMSSQSDSCISMROSGCLPLLIQLL 358
 Qy 361 HGNKDVSLLGNSRSGKEARASAAALHNIHSQPDKGRREIRVHLLEQIRACETC 420
 Db 359 HGNKDVSLLGNSRSGKEARASAAALHNIHSQPDKGRREIRVHLLEQIRACETC 418
 Qy 421 WEQEAHEPGMDQKNMPAPVEHQICPAVCVLMKLSFDEEHRHANNELGGQAIALLQ 480
 Db 419 WEQEAHEPGMDQKNMPAPVEHQICPAVCVLMKLSFDEEHRHANNELGGQAIALLQ 478
 Qy 481 VDCMYGLTNDHYSTILRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQKSEEDL 540
 Db 479 VDCMYGLTNDHYSTILRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQKSEEDL 538
 Qy 541 QQVIASVLRNLSWRADVNSKTLREYGVSKALMECALEVKESTLKSLSALWNLSAHCT 600
 Db 539 QQVIASVLRNLSWRADVNSKTLREYGVSKALMECALEVKESTLKSLSALWNLSAHCT 598
 Qy 601 ENKADICAVDGAFLVGLTYRQNTLAIIESGGILRNVSLLIATNEDHROIIRENN 660
 Db 599 ENKADICAVDGAFLVGLTYRQNTLAIIESGGILRNVSLLIATNEDHROIIRENN 658
 Qy 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKQDALWDMGAVSMLKNLIHSHKHMIAM 720
 Db 659 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKQDALWDMGAVSMLKNLIHSHKHMIAM 718
 Qy 721 GSAALARNLMANRPAPKYKQANINSPGSSILPSLHVRRKQKALEAELDAQHLSETPDNIDLS 780
 Db 719 GSAALARNLMANRPAPKYKQANINSPGSSILPSLHVRRKQKALEAELDAQHLSETPDNIDLS 778
 Qy 781 PKASHRSKOPHKQSLYGVYVDNTRDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSSRGS 840
 Db 779 PKASHRSKOPHKQSLYGVYVDNTRDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSSRGS 838
 Qy 841 LDSRSSEKDSLRERERGIGLGNTHPATENPGTSKRGLOITSTTAQIAKYMEVSVSAHTS 900

RESULT 2
 I49505
 adenomatous polyposis coli protein - mouse
 N:Alternate names: APC
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 C:Accession: I49505

Db 839 LSSRSSEKDRSLRERIGIGLSAYHPTTENAGTSSKRGQLQITTTAAQIAKVMREVSALHTS 898
 Qy 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSTHNTYNTFKSENRRCTSCMPYAKLEYKRSS 960
 Db 899 QDRSSASTTEFFCVAADRRAARSSASHSTHNTYNTFKSENRRCTSCMPYAKLEYKRSS 958
 Qy 961 NDSLSNVSSSDGKGKGMKPSIYESSEDESXPCSYGVYPADLAHKIHSANHMDDNCE 1020
 Db 959 NDSLSNVSSSDGKGKGMKPSIYESSEDESXPCSYGVYPADLAHKIHSANHMDDNCE 1018
 Qy 1021 LDTPIYSLKYDEQLNSGRQSPQNERWARPXHIIEDEIKOSEQRQSNQSTTPYVYTE 1080
 Db 1019 LDTPIYSLKYDEQLNSGRQSPQNERWARPXHIIEDEIKOSEQRQSNQSTTPYVYTE 1078
 Qy 1081 STDKHLKFPQHPGQCECVSPYSRGANGSETNVCNHCINQNVQSICQEDDDYEDDKP 1140
 Db 1079 NTDDKHLKFPQHPGQCECVSPYSRGANGSETNVCNHCINQNVQSICQEDDDYEDDKP 1138
 Qy 1141 TNSERYSEEEQH-EEBERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSPFSK 1199
 Db 1139 TNSERYSEEEQH-EEBERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSPFSK 1198
 Qy 1200 SSGQSKTHMSSSENTSTPSNAKRONLHPSSAQSRGQPOKAACTKVSSINQETI 1259
 Db 1199 NSSAQSTKPEHLSPSENTAVPPSPNAKRONLHPSSAQ-RNGQTQKGTTCVKPSPINQETI 1257
 Qy 1260 QTCVDETPICFRKCSLSLSAEDEIGCNOTTOEADSAANTLOIAEIKKIGTSAEDP 1319
 Db 1258 QTCVDETPICFRKCSLSLSAEDEIGCNOTTOEADSAANTLOIAEIKKIGTSAEDP 1317
 Qy 1320 VSEPAVSGHPRTKSRLOSGSLSSSARH-KAVEFSSGAKSPSKSGAQTPKSPPEHYVQ 1378
 Db 1318 ATEVPAVSGHPRTKSRLOSGSLSSSARH-KAVEFSSGAKSPSKSGAQTPKSPPEHYVQ 1377
 Qy 1379 ETLPMFSRCTSVSLDSFESRSTASSVQSEPCGMVSGIISPSDLPSDQGMTPSRSKT 1438
 Db 1378 ETLPMFSRCTSVSLDSFESRSTASSVQSEPCGMVSGIISPSDLPSDQGMTPSRSKT 1437
 Qy 1439 PPPPQTAQTKREVPKNAKPTAEKRESGPKQAANAQVQVLPDADTLHFATSTPD 1498
 Db 1438 PPPPQTAQTKREVPKNAKPTAEKRESGPKQAANAQVQVLPDADTLHFATSTPD 1497
 Qy 1499 GFSCSSLSALSDEPFIQKVELRIMPVQENDNGNETESQPKSENQEKAEKTIID 1558
 Db 1498 GFSCSSLSALSDEPFIQKVELRIMPVQENDNGNETESQPKSENQEKAEKTIID 1556
 Qy 1559 SEKDLDDSDDDIEILEECIIISAMPTKSRKAKPAQATSKLPPPVARKPSQLPYKLL 1618
 Db 1557 SEKDLDDSDDDIEILEECIIISAMPTKSRKAKPAQATSKLPPPVARKPSQLPYKLL 1616
 Qy 1619 PSQNRLOPKHVSFTPGDDMPRVYCVETGPIINFSTATSLSDTITSPPNELAAAGVRGG 1678
 Db 1617 PAQNRLOPKHVSFTPGDDMPRVYCVETGPIINFSTATSLSDTITSPPNELAAAGVRAG 1676
 Qy 1679 AQSGEFKRTDITEGRSTDEAGCTTSVTIPELDDNKAEEGDILAEICINSAMPKGSK 1738
 Db 1677 IQSGEFKRTDITEGRSTDEAGCTTSVTIPELDDNKAEEGDILAEICINSAMPKGSK 1736
 Qy 1739 KPRVKKIMDQVQASASAPNKNOLDGKKKPTSPVKPIFQNTYRTRVRKNADSKN 1798
 Db 1737 KPRVKKIMDQVQASASAPNKNOLDGKKKPTSPVKPIFQNTYRTRVRKNADSKN 1796
 Qy 1799 LNAERYFSNOKSKONLKNNSKDFNDKLPNNEDVRGSPFADSPHHYTPIEGTVCFSR 1858
 Db 1797 VNTETFSNOKSKONLKNNSKDFNDKLPNNEDVRGSPFADSPHHYTPIEGTVCFSR 1856
 Qy 1859 NDSLSLDDDDDDVLSREKAEELRKAENKESAEKVTSTELTSSNQSSANKTQAIAKQPI 1918
 Db 1857 NDSLSLDDDDDDVLSREKAEELRKAENKESAEKVTSTELTSSNQSSANKTQAIAKQPI 1916
 Qy 1919 NRCQPKILQKSTPQSSKQIDPCRAATDEKLOFAIENTVPCFSHNSLSLSLDDIOE 1978
 Db 1917 NRAQSKPVLQKSTPQSSKQIDPCRAATDEKLOFAIENTVPCFSHNSLSLSLDDIOE 1976

Qy 1979 -NNKNEPIKETPPDSQGEPSKPOAGYAPKSFHVEDTPVCFSRNSSLSLSLSDSEDD 2037
 Db 1977 NNNKNEPIKETPPDSQGEPSKPOAGYAPKSFHVEDTPVCFSRNSSLSLSLSDSEDD 2036
 Qy 2038 LQBCISSAMPKKKPSRLKGDNEKHSPLNNGGITLGEDTLDLKDIORPDSEHGLSPDSE 2097
 Db 2037 LQBCISSAMPKKKPSRLKGDNEKHSPLNNGGITLGEDTLDLKDIORPDSEHGLSPDSE 2096
 Qy 2098 NFDKALQEGANSIVSSLHQ-AAAAACLRSQASDSISLSKSGISLGSFPHLTPDOEE 2156
 Db 2097 NFDKALQEGANSIVSSLHQ-AAAAACLRSQASDSISLSKSGISLGSFPHLTPDOEE 2156
 Qy 2157 KPFTSNKGRPLKPEKSTLETKIESEKIGKGVKYSLLITGKVRNSEISGQWQKOP 2216
 Db 2157 KPFTSNKGRPLKPEKSTLETKIESEKIGKGVKYSLLITGKVRNSEISGQWQKOP 2216
 Qy 2217 LQANPISIRGRMTIHPGVNRSSSTSPVSKGPPKLPKTPASKSPSEGGTATTSPRGAKP 2276
 Db 2217 LQANPISIRGRMTIHPGVNRSSSTSPVSKGPPKLPKTPASKSPSEGGTATTSPRGAKP 2276
 Qy 2277 SVKSELSPVARQTSQIGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPRNGI 2336
 Db 2277 SVKSELSPVARQTSQIGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPRNGI 2336
 Qy 2337 SPPNKLQOLPRTSPSTASTKSGSGKVSYSYTPGRQMSQQNLTKQCLSKNASSIPRSES 2396
 Db 2337 SPPNKLQOLPRTSPSTASTKSGSGKVSYSYTPGRQMSQQNLTKQCLSKNASSIPRSES 2396
 Qy 2397 ASKGLNOMWNGNANKVEELSRMSSTKSGSESDRSEBPVLRQSTFIKEAPSPFLRKL 2456
 Db 2397 ASKGLNOMWNGNANKVEELSRMSSTKSGSESDRSEBPVLRQSTFIKEAPSPFLRKL 2456
 Qy 2457 BESAFSLSLSPSSRSPASPTRSQATPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIE 2516
 Db 2457 BESAFSLSLSPSSRSPASPTRSQATPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIE 2516
 Qy 2517 YNDGRPAKRDHDIARSHESPSRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSILSA 2576
 Db 2517 YNDGRPAKRDHDIARSHESPSRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSILSA 2576
 Qy 2577 SSESSEKAKSEDEKHVNSISGTQSKENQVSAKGTWKIKENEFSPNTNSQTVSSGATN 2636
 Db 2577 SSESSEKAKSEDEKHVNSISGTQSKENQVSAKGTWKIKENEFSPNTNSQTVSSGATN 2636
 Qy 2637 GAESKTLIYQMAFAVSTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKSSSIKD 2696
 Db 2637 GAESKTLIYQMAFAVSTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKSSSIKD 2696
 Qy 2697 SKDNQ---AKQTVNGSVPMRTVGLNRLNSFTQVDAPDQKQTEIKQGNPNVPVSETNE 2753
 Db 2697 SKDNQ---AKQTVNGSVPMRTVGLNRLNSFTQVDAPDQKQTEIKQGNPNVPVSETNE 2753
 Qy 2754 SSIVERTPFSSSSSSKHSPPSGTVAARVTFNTPNPRKSSADSTARPSPQITPVNNNT 2813
 Db 2754 SSIVERTPFSSSSSSKHSPPSGTVAARVTFNTPNPRKSSADSTARPSPQITPVNNNT 2813
 Qy 2814 KRRDSKTDSTESSGTQPKHSGSYLVTSV 2843
 Db 2814 KRRDSKTDSTESSGTQPKHSGSYLVTSV 2843
 Qy 2816 KRRDSKTDSTESSGTQPKHSGSYLVTSV 2845
 Db 2816 KRRDSKTDSTESSGTQPKHSGSYLVTSV 2845

RESULT 3

T30258

adenomatous polyposis coli protein 2 - mouse

N:Alternate names: APC2 protein

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000

C:Accession: T30258

R:van Es, J. H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Miles, A.; Kuipers, J.

Curr. Biol. 9: 105-108, 1999

A:Title: Identification of APC2, a homologue of the adenomatous polyposis coli tumour sui

A:Reference number: Z20796; MUID:99147086; PMID:10021369.

Qy	875	KGCLQISITTAQIAKWEVESAHTSOEDRSSGSGTTELCVTDERNALRSSAAHTSNT	934
Db	834	KAKAKALAVARIDRUVEDISAHTSSDSDFSUSS	887
Qy	935	YNPT--KSENSNRCTCMPYAKLEYKRESSNDLSNVSSDYGKRGOMKPIESYEDDES	992
Db	888	RGTEGGRREAGSRAHPLLRKAAHTSLSDNSLNGSGTSDGYCTREHWP	936
Qy	993	KFCSYQOYPADLAHKIHSANHMDNDGELDTPINYSLKYSDEQLNSGRQSPSQNERWAP	1052
Db	937	--CPLAAL--AHRDD	956
Qy	1053	KHIIEDEIKQSEORQSRNOSTTYPVYTESTDDXHLKFQPHFGQOECVSPYRSRGANGSET	1112
Db	957	RLDLDPLSRAELPARDTAAATDARVET	1002
Qy	1113	NRVGSNHNQNVQSQLOEDDEDYDKPTNYSERYEEBOHEEERPTNYSIKYNEKPH	1172
Db	1003	-----GVRPLVGPCTS	1013
Qy	1173	VDQPIDSLKYATDIPSSOKQSPFSKSSGQSSKTEHMSSEN*STPSSNAKRONQLH	1232
Db	1014	-----PGARKQAW	1021
Qy	1233	PSSAQSRSGOPOKAA*CKVSSINQETIQYCVBEDTPICFSRCSLSLSLSAEBDEIGNQOT	1292
Db	1022	-IPADSLSKVPEKLVASPL-PIASKVLQKLVAQDGPM*SLRCSLSLSLSSTGHAVPSQAE	1079
Qy	1293	TQADSANVLTQIAEIKETGTSAE	1340
Db	1080	NLSDSS-----LEGLEEAGPGAELGRAWRAGSGTSLPVS-IPA-----PQGRSR	1125
Qy	1341	SLSSGARHKAVFSGSAKSPKSGAQTPKSPPEHYVQETPLMFSCRCTSVSSLDSFESRS	1400
Db	1126	GLGVEDA-----TPSSSENCVQETPLVLSRCSSVSSLSGSPESRS	1165
Qy	1401	IASVSGSEPCSGMVSGIISPSDLPDSPGQTMPPSRSKTPTPPPTQATXR	1454
Db	1166	IASSIPDPCSLGSGTVSPSELPSDPSGQTMPPSRSKTPTPPAPGQPE*TSQFSLQWESYVK	1225
Qy	1455	NKAPTAKEBESGKQAAVNAQVQVLPDADTLLHFATSTPDGSCSSLSALSDEP	1514
Db	1226	RFIDIDACRCCOPPELDAGSV	1273
Qy	1515	FIQKVELRIMPPV-QENDNGNETESEQPKESNENQEKAEKTI*DSK-----DLLDSDDD	1570
Db	1274	YVQDVELRLRPACPERAVGG-----GGHRRRDEAASRLDGPAPAGSARSATDK	1324
Qy	1571	DIILIECIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPSONRLQPOKHV	1630
Db	1325	ELBALRECLGAAMPARLK-----VASALVP--GRSLUPVPYMLVPAPAR	1368
Qy	1631	SFTPGDDMPRVYCVGEGTPIINFSTATSLDUTIESPPNELAAGVGVRGGAOSGEFEKRTDI	1690
Db	1369	---GDD-SCTDSAGTIPVNFSSAASLSD*ETLQGPSRDKPAGPGR-----QK-----	1411
Qy	1691	PTGSRSTDEAQGKTSVTIPELDDNKAEBEGDILAEICINAMPKGSHKHFRVKKINDQV	1750
Db	1412	PT-GRAAPARQ-----TSRHP-----	1427
Qy	1751	QQASASSAPKNQLOGKKKPTSPVKPIPQNT*EY-RTRVRKNADSKNNINAEVRFSDNK	1809
Db	1428	KAAGAKS-----TEHTGPCENRAGLELPLSRPQASRSNR	1463
Qy	1810	DSKKQNLKNNSKDFNDKLPNNEDVRGSAFADSPHHYTP*IEGTPYCFSRNDSJSSLDFDD	1869
Db	1464	DSSCOT-----RTRDGGALQSLCLTPTTEAVYCF-----YDS	1496
Qy	1870	DDVDLS-----BEKALRKAENKSEAKVSHTSLT*SNQOSANKTQATAKOPIN--R	1920
Db	1497	DEEPPATAPPRRASATPRALKREKPAGRK-----ETPSRAQPATILPVR	1541
Qy	1921	GQPKPILQKSTFPQSSKQIPDRGAATDEKQLQFAIENTVCFSHNSLSLSLSDIQENN	1980

770 -----ENATKTSY----- 778
 Db 1110 SETRVGSHGINQVSQLCOEDDYEDDKPTNYISYSEBEEHBERPTNYSIKNEE 1169
 Db 779 -----QETDL--DQPTDFSLYAE-----NQIESLDLISGPAGGQ 811
 Qy 1170 KRHDQPIDYSLKYATDIP--SSQKQSFSSKSSGQSKTEH-----MSSSENTSPSS 1223
 Db 812 KSTITPP-----AETVPEKSEGEIILLIDDSVKCYQTEDPYVISNAASVTDLRVAA 864
 Qy 1224 NAKQNLQHPG--SAQSRGQOPQ---KAATKQVSSINOETIQTVCYVEDTPICFSRCSLS 1278
 Db 865 KADREAEVKPEVREVTKEGAPKPLKLSQCGSGSYTPKPIKYCEBGTFCYFSRYDSLS 924
 Qy 1279 SLSSAEDEIG--CNQTTQEASANTLQABIEKIGTSAEDPVSEVFAVSHCHPRTKSSRL 1337
 Db 925 SL-----DESGKANAQIVGTD-----ADIKPKLEKEEESQPAEQVLTKPPTQANSAL 973
 Qy 1338 QGSSLSSESARHKAVFSSGAKSPKSGAQTPKSPPEHYVOETPLMFSTRCTSVSSLDSE 1397
 Db 974 -----ETPLMFSTRSSMDSLVHDP 992
 Qy 1398 SRSTA-----SSVOSEPCSGMVSGIISPSDLPSDPTMPSPRSKTPPPPTQACTKEV 1452
 Db 993 DVDVANDCKSSVSD--FSRLASGVISPEIIPDPTQSMPOS----- 1033
 Qy 1453 PKNAPTAEKTESGPKOAAVNAQVQVLPDADTLHFATESTPDGFGSSSLSALSLD 1512
 Db 1034 PRNSVAGSGQVNDSPVIVFASIQPLRSVFE--DDLSEFVETPAQFSTATSLNLSI-- 1091
 Qy 1513 BPFTQKQVELRIMPVQENDNGNE-----TESHQPESNENQKE--AEXTID 1558
 Db 1092 -----VDDEKAPAVMTEDNEDELLANCINMGQRKTEAVKSTVYNSEVDVAEETIR 1144
 Qy 1559 S-----EKD-----LLDD--SDDDDI-- 1572
 Db 1145 SYCTEDTPALLSKVPSNTLSVISMSTDPKDATAGQAQMAHQSLDDVSNASDCOGAS 1204
 Qy 1573 -EILEECIIISAMPTKSRKAKKPAQASKLPPPVARK--PSQLPVYKLLPSONRLQPOKHV 1630
 Db 1205 GHLLQOCIRDM-----KKPLGEATSDPIAMLRGNGELPGY--LPS----- 1244
 Qy 1631 SFTPGDDMPRYVYEGTPINSTATSLDITIES-----PPNELAAGVRGQSGGSEFEX 1686
 Db 1245 -----ADENMK--FLVEDSPCNFVSGSLNLTGVGSLGPAVOLKETE--PSSADQNPENKR 1298
 Qy 1687 RDTIPTGRTDEAQGGKTSTVITPELDDNKABEGDILAEQ----- 1727
 Db 1299 SLANRSKRRPPHWDQDSSLSLSDSEDDTNLLSQATAAGCNRPKSNLGFSSNGKRSSSL 1358
 Qy 1728 -----INSAMPKGSHPKPRVKINDVOQOASASSAPWQO-----LDG--KK 1769
 Db 1359 SSSQPIAINAATSASSLNSAMTVKRSQQESYSYSDSDSDNOSKSLFELCILKGMKYT 1418
 Qy 1770 KKP-----TSPVKPIPTQNTYRT-----RVKNADSKNLAER 1803
 Db 1419 KEPGARAQMOEQPIVGSVSSVQSPSLKQFDSLVPQLPSSQVVRQHHHHHHHRRER 1478
 Qy 1804 VFSNKKSK--KONLKN--SKDFNDKLPNNEDVRGSPAFPSPHYTYPIETGTYCFGRND 1860
 Db 1479 ---ERKDEKLQECINTGISKKIN--AVPKNV--LATAAALPECH--PMAAT-----T 1523
 Qy 1861 SLSSLDFFDDVDLSREKAEKLRKAKENKESAKVTSHTLNSQOSANKTOA--IAKOPI-- 1918
 Db 1524 SASALSTAAPVE-----QKAH-----ATSNPQOSSTHPSHILPNPIDATVTDTRGPA 1577
 Qy 1919 -NRGQPKPILOKQSTFPOSSKDIPRGAATDEKLQNAFIEN-----TPVCF----- 1963
 Db 1578 PNOGNGNSONGLET--ATGSKOLDSEDSDESQSFIMETWRLDSALNETCISGASEK 1636
 Qy 1964 -----SHNSLSLSLSDIOENKNNKENEBIKETEPDS 1995
 Db 1637 HKOPDLMLKSVRLTMEFTVSAEQLRSSSHHSS-----NSHKNNKSNNTWNTESTCPND 1691

RESULT 5

T13564
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
 N;Alternate names: Hypothetical protein EG:4984.1
 C;Species: Drosophila melanogaster
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C;Accession: T13564
 R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
 submitted to the EMBL Data Library, April 1999
 A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A;Reference number: Z17689

Qy 1996 QGEPKPOASGYAPKSFHVEDTPVCFSRNSSLSLSID---SEDDLQECISAMPKKKK 2052
 Db 1692 VFPFVSQTA-----PV-----LASLSDDEDATEARSLHELIIITPTNEQQ 1732
 Qy 2053 PRLKGDNEK-----HSPRNMGGILGEDLTLDIKIDRPDSEHGLSPDSENFQWAKIQEGA 2108
 Db 1733 PESLEGETDLVNGHADSYSYGSSGGLNFOL--GGQVQVAGVRLRPEORLLFFNGTSASIMT 1789
 Qy 2109 NIIVSSLHAAAAACLSROASSDS--SLSL-----KSGI--SLGSPF--HLTP 2152
 Db 1790 NMTIATFAEARALAEALLQPAATDDTTMTFSLNSLDLDNIRPPSGMESLNCYQDHSOF 1849
 Qy 2153 DQEBKFTNKGPRIKPGCKSTLETKIESKGIKGGKVKYKSLITKVASNSEISQ 2212
 Db 1850 SSRLQAMPS--KSPRPARKVFPANLVARRALHLAG-----SAESVNSSCNLLDN 1897
 Qy 2213 MKQP-----LOANMPSISRGRTMIHIGVNRNSSSTSPVSKGPPPL----- 2253
 Db 1898 IKPPLMLDELDSMISVDISQSEVADGEQDCSMATTISVSNYETAAACDDQTMVLQSCFD 1957
 Qy 2254 -----KTPASKSPSEGQTATTSPRGAKSVKSEL-----SPVAR 2287
 Db 1958 EDEDATMNDYSAESTPKHGSTPSNRRSLTPDKRRLTKDRFTYTIATSCMEAPEAN 2017
 Qy 2288 QTSQI-----GGSKAPSRSGSRDSTPSRPAQOPLSRPI----- 2321
 Db 2018 ETLOIEIVEAAVPAVTPSPRANGRRGSAERYKTOLIECPALIOQPDDCPSEQLSIR 2077
 Qy 2322 -----CSP--GRNSISPRNG-----ISPENKLS 2343
 Db 2078 AMWQFTFTDINIGHSQTCSTHDPEDAGSEPCDQNSCTESCDGQBPDLPPPSIV 2137
 Qy 2344 QL-----PRTSPSPASTKSSGSGHMSYTSRQMSQQNLTKQTGLSKNASSIPRSESA 2397
 Db 2138 DLRTSVVKPTTLEPATAVKLVGRKKPAVSPVSMQSRN----- 2177
 Qy 2398 SKGLNOMNGANGANKVELSRMSTKS-----SGESDRSERPV-----LVQSTF 2443
 Db 2178 -----SNNAPSKKKTLSPTIAKSLVPGSGVRLPAKKKTPPPPEPAPARLEQGT 2230
 Qy 2444 IKEAPSPTLRRLKEASPELSPSRRPASPRSQATPVLSPSLPDMSLSTHSSVQAG 2503
 Db 2231 VKD-----EPNSNVQVFPVE--TKDAQTSPIHRA----- 2258
 Qy 2504 WRKLPNLSPTEYNDGRPAKHDIARSHESPSRLPINRSGTWKREHSHSSSLPRVST 2563
 Db 2259 -SKLPTKKGTA--SGGSPSK-----AGSPKRIPLAFA--RMTQPRANTSRLAA 2303
 Qy 2564 WRGTGSSSILSASSSESEKAKSEDEKHVNSIGTKQSKENOVSAK--GTWRKIKENEFS 2621
 Db 2304 GKPEAASRVVSGRVSTTPPSRNSNLNGSSAAAAAKINQAQSIANIKRVDKATK 2363
 Qy 2622 PTMS---TSQTVSSGATNGAESKTLIYQMAPAVSKTEDYVVRIEDCPINPRSGRPTGN 2678
 Db 2364 QSSSLRTQKTSNMLNANGTKPTLLR-----SSTFD-----N 2397
 Qy 2679 TPVIDSVSEK 2689
 Db 2398 TPSTAGGVASK 2408

A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DDBU
A;Molecule type: DNA

A;Residues: 1-5327 <SPA>

A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1

C;Genetics:

A;Cross-references: FlyBase:FBgn0025392

A;Introns: 24/2; 52/3; 104/3; 179/1; 1669/3; 2566/1; 4798/3; 5272/1

A;Note: EG:4924.1

C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.5%; Score 649; DB 2; Length 5327;

Best Local Similarity 18.2%; Pred. No. 3.1e-18;

Matches 614; Conservative 521; Mismatches 1263; Indels 968; Gaps 129;

QY 15 ALKWNENLRLQELE-DNSNHLTKLETEAGNMKEVLKQLQGSIEDAMASSGQIDLLERLK 73
D 15 ALKWNENLRLQELE-DNSNHLTKLETEAGNMKEVLKQLQGSIEDAMASSGQIDLLERLK 73
D 819 AKKLQDLTASQELDAEKQRELDLKEE---QEVVREIEAVFSRDMKQEQQIQKAEUR 874
QY 74 EMLN-----DSSNPPGVK-----LRKMSLRSYSGREGSV-----SSRGECSPPVPMGSFPRR 121
D 74 EMLN-----DSSNPPGVK-----LRKMSLRSYSGREGSV-----SSRGECSPPVPMGSFPRR 121
D 875 EMPAEGTGGENPDEEYELIIEKEVEQY--TEDSIVEQSSMTKE----- 920
QY 122 GFVNGSRESTGYLEELKESLILLADLKEEKEKWMYLAOLNLTAKRIDSLPLT----- 175
D 122 GFVNGSRESTGYLEELKESLILLADLKEEKEKWMYLAOLNLTAKRIDSLPLT----- 175
D 921 -----EEIQKHQR-----DSQSEKKEKSAEIEAIAKVAEAEKARL 961
QY 176 ENFSLOTMTTRQLEYEARQIRVAMEEQLGTODMEK-RAQRRIAR----- 220
D 176 ENFSLOTMTTRQLEYEARQIRVAMEEQLGTODMEK-RAQRRIAR----- 220
D 962 EGASARQDESELDVEPEQSKIAEQDIIATANDIAKSTRTEQLAKPAEELSSTPTEK 1021
QY 221 -----LQOLEK 226
D 221 -----LQOLEK 226
D 1022 LSKKTSKDDQIGAPVDVLPVNLQESLPEEFKSFATIESGATTAPLTPDERIPIDQIKE 1081
QY 227 DILRIQLLOSQATEAB-----RSSQKHQ-----TGSHDAERQNEQGVGEINM 271
D 227 DILRIQLLOSQATEAB-----RSSQKHQ-----TGSHDAERQNEQGVGEINM 271
D 1082 DLVIEKYVKEETKEAEIATVATQTLPEAAPLAIDTILASATKADKADANAALGE--L 1139
QY 272 ATSGNQGSTRMDHETASVLSSSTHAPRLTSHLGTKEVMYVSL-----LSMLGTHDK 327
D 272 ATSGNQGSTRMDHETASVLSSSTHAPRLTSHLGTKEVMYVSL-----LSMLGTHDK 327
D 1140 PDSGE-----RVLPMTKTFEACQNLRLRDVIKTPDEVADLPVHEEADLGLYEK 1186
QY 328 DMSRTILLAMSSQDSCISMROS-----GCLPLLIQLLHGNDKDS----- 367
D 328 DMSRTILLAMSSQDSCISMROS-----GCLPLLIQLLHGNDKDS----- 367
D 1187 DSODANAKSISHKEESAKEKETDDEKENKVGIELGDEPNKVDISHVLLKESVOEVAEK 1246
QY 368 VLLGNRSGKEAPARASALHNIHQPDDKGRREIRVLHLEQIRAYCETCWQEAH 427
D 368 VLLGNRSGKEAPARASALHNIHQPDDKGRREIRVLHLEQIRAYCETCWQEAH 427
D 1247 VVIETTVKKQBEIIEATVITQNEQD-----LMEQVKD-----KEEH 1286
QY 428 EPQWD-----ODKNPMPAPVEH-----QICPAVCLMKLSFDEEHRHAMN 467
D 428 EPQWD-----ODKNPMPAPVEH-----QICPAVCLMKLSFDEEHRHAMN 467
D 1287 EQIESGIITEKEAKSASTPEKETSDITSDELPAQLADPTTTPPKSAKDRE----- 1340
QY 468 ELGGLQA-----IALLQVDCWYGLTNDHYSITLRRYAGMALTNLTFGDVANKATILCMKG 524
D 468 ELGGLQA-----IALLQVDCWYGLTNDHYSITLRRYAGMALTNLTFGDVANKATILCMKG 524
D 1341 DTGSIESPPTIEAIEVE----- 1358
QY 525 CMRALVAQLKSESDIQQVITASVLNLSMRADVNSKKTILREVGSVKALMECALEVY-KES 583
D 525 CMRALVAQLKSESDIQQVITASVLNLSMRADVNSKKTILREVGSVKALMECALEVY-KES 583
D 1359 -----VQAKQEAQ--KPDAPAEALIKTEKSPASKETSPPESATSGSVKEDTEQTKSKS 1410
QY 584 TLKSVLSALNLSAHTENKADICAVDGAFLVGLTYRSQNTLAIIESGGGILRNVS 643
D 584 TLKSVLSALNLSAHTENKADICAVDGAFLVGLTYRSQNTLAIIESGGGILRNVS 643
D 1411 PVPD-----RPESEAKDKSPASGEAS--RPESVAESVKDEAGKAESRRE 1454
QY 644 SLIATNEDHROI-----LRENNCLOTLLHLSLITVSNACGLTNLNLNLSARNPKDQEAALWD 700
D 644 SLIATNEDHROI-----LRENNCLOTLLHLSLITVSNACGLTNLNLNLSARNPKDQEAALWD 700
D 1455 STAKTHKDESSLDKAEQSRRESLAEISIKPSGGIDKSA-----LASKAESRPESVTD 1508
QY 701 MGAVSMKLNLIHSHKRWIANG-SAAALRNLMANRRPAKYKDNANIMPGSSLSLHYVRKOKA 759
D 701 MGAVSMKLNLIHSHKRWIANG-SAAALRNLMANRRPAKYKDNANIMPGSSLSLHYVRKOKA 759

Db 1509 -----KSKEPSRRESIAESLKAESTKDEKSAAPPK-----EASRPGSVVSVKDETEKS 1557
QY 760 LEALDAQHILSETFDNDNLSPKASHRSKORHKOSLYGVDVFTNRRHDDNRSNFNTGNM 819
Db 1558 KEP-----SRRESIAESAKPIEFREVRPSPSVI--DGIKDESAKPESRD----- 1601
QY 820 TVLSPYLNTTVLPSSSSSR--GSLDSSRSSEKORSKORLERIGIGLNHPA-----TENPGT 872
D 1602 ---SP-----LASKEASRPESVLESVKDEBPXSTKESRRESVAESPKADSTKDEKSPLT 1652
QY 873 SSXRGLOISITAAQIAKMBEVSALHTSQSDRSGSGSTTELHCTVDERNALRRSSAAHTHS 932
D 1653 SK-----DISPESAVENM-DAPFKETSRPESAVGS-----MKDESMSKEPS 1694
QY 933 NTYNFTSESNRTCTMPYAKLEYKRRSSNDLSNVS---SSDGYGKRQGMKPSIESYSED 989
D 1695 RRESVKDGAQSRSTRSPASVAESAKDGADLKLRSPESTTQSKAAGSIKDEKSPLEASE 1754
QY 990 DESKFCSYGOYPADLAHAKIHSANHMD-----NOCGLDTPINYSIKYSDEQLNS----- 1038
D 1755 EASRPASVAVSVKDEAKESKEESRRESVAESKPLPSKEASRPASVAESIKDEAKESKEES 1814
QY 1039 -----GRQSPSQNERWARPKHI---TEDEIKQSEQ---RQSRNQSTTYP-----VYT 1079
D 1815 RRESVAESKPLPSKEASRPASVAESIKDEAKESKEESRRESVAESKPLPSKEASRPASVA 1874
QY 1080 ESTDDKHLKTPHFGQECV-----SPYRSGANGSETNRVGSNHNQVNSQSLCQE 1132
D 1875 ESIDEAKESKEE--SRRESVAESKPLPSKEASRPASVAESIKDEAKESKEESRRESVAEK 1933
QY 1133 D---DYDDDKPTNYSEYSEE-EQHEERPTNYSIKYNEEKHRVQPIDYSLKYATDIP 1188
D 1934 SPLPSKASRPASVAESIKDEAKESKEESRRESVAESKPLPSKEASRP-----ASVA 1985
QY 1189 SSQKQSFYKSGSGQSKTEHMSSESSSENTSTPSSNA-----KRONLHPSSQAQRS 1240
D 1986 ESIDEAKESKEESRRESVAESKPLPSKEASRPASVAESIKDEAKESKEESRRESVAES 2045
QY 1241 GQPKAATCKVS-----SINGETIOTYCEDTPICFSCSSLSLS--SAEDRIGC 1289
D 2046 PLPSKEASRPASVAESIKDEAKESKEESRRESVAESKPLPSKEASRPASVAESVKDE--A 2103
QY 1290 NQTTQE-----ADSANTLQI---AEIKEIGITRSEAEDPVSEVPVAVSQHPRTKSRLQG 1339
D 2104 DKSKEESRRESMAESGKAQSIKGDQPLKEVSRPESVAESVKDDPVKSEPSRRES--VA 2161
QY 1340 SLSSESARHKAVEFFSGAKSPKS-CAQTPKPPPEHYQETPLMPSRCTSVSSLSOSFES 1398
D 2162 GSVTADSGARD-----QSPLESKASRPESVDSVKDEAKESRRES----- 2204
QY 1399 RSIASSVQSPFCGSMVSGIISPDLDP-SFGQTMPP--SRSKT-----PPPPQTAQTKR 1450
D 2205 -----KTESVIPPKAQKODKSPKEVLQPVSMVTETIREDAQPMKPSQAESRR 2250
QY 1451 E-----VPMK-KAPTAEKRESGPKQAA--VNAAVQVQVLPDADTLHLHATSTPD 1498
D 2251 ESTAESIKASPRDEKSPKASKEASRPASVAESIKYDLDKPKIHKDSTHEHRSRESLED 2310
QY 1499 GFSCSSLS----- 1507
D 2311 KSAVTSEKSVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGFVAETVSSPIEATMEF 2370
QY 1508 -----ALSLD-----EPFIQKDELIRIM 1525
D 2371 SKIEVVEKSLALSLOGSGGKLQTDSPVDVAEGDFSHAVASVSTVTTLTTPKPAELAQI 2430
QY 1526 -----PPVOENDNGNETESEQPKSENENQKEAKETIDSEKDLDDSD 1569
D 2431 GAAKTVSSPLDEALRTPSAPEHISRADSPACASEIASODKSPQVLKSSSRPAWVAESK 2490
QY 1570 DDTEILEECIISAMPTKSSSKAKKPAQATSKLPPPVARKPSQPLVYKL----- 1617
D 2491 DDAQLKSSVEDLRSFVASTETISRPAAGETASSPIEAPKDFABEQAEKAVPLUTIEL 2550

Qy		1668	ELAAEGVGEVGAAGCAQSGEPEKRDTP-TEGRSTDEAOGGKTSSVTIPB-----LDDNKABE	1720
Db		1081	--TEASSTGGGSTSPNPGQSTSPSTGATSPGSGSGLTILTSISPSPQSSITIGSQGST	1138
Qy		1721	GDILAECINAMPKGSHKKPRFKIMDVOQAASASSAPNKNQLDKKKKPTSPVKPIP	1780
Db		1139	SPVSTSGDMTSQGSTQIPGSGTGTVTPGSGSGSTSGEITSQGTQTPRSSJLSTSP	1198
Qy		1781	QNTERTVRKXADSKMNLNAERVPNDKDSKONLKNNKDFNDKLPNNEEDRVRSFAF	1840
Db		1199	AIS-----TSTQQSVSTNSPGSTVTPQST---VRGSTS	1229
Qy		1841	DGPHPHTPIEGTPYCFSRNDSLSSLDFFDDDDVLSPREKAELRAKENKESEAKVTGHTEL	1900
Db		1230	GSTVTGTSGES-----STSGSSSATSLSSSPV	1258
Qy		1901	TNQGSANKTOIAIKQPINRGOPKPILOKQSTPPQSSKDIDREGAATDEKLQNFALENTP	1960
Db		1259	PSTSOSPNSTSGSSTP--TFNP--SQSTSPVVS-----TTTGEMTHSGSTQTPT	1303
Qy		1961	VCFHSNSSLSLSDIOENNNKENEPIKEETEPDSQCE-PSKPQAGYAPKSPHEVEDTPV	2019
Db		1304	STI--GSTVQTSTVGCSNSSGSTVTIGSEASTSGSSPKTSPSSIPIPVTTSPPISTTF	1361
Qy		2020	CFSRN-SLSLSLSIDSDLLQECISSAMPKKKPRLKGDNEKHGPRNMWGGILGEDJTL	2078
Db		1362	ASSTSGSTISDVSSVSTTSLAP--LSSSLPSTV-----	1392
Qy		2079	DLKDORPSEHCLSPDSENFDWKATQEGANSIVSSLHQAAAAACLSRQASSDSDSJLSL	2138
Db		1393	-----PSTQSFSST-----EGSSKASSSPVPQSSTPTPNPTGESSTL-L	1435
Qy		2139	XGSI LSGSPFHILTPOQ-----EBKFPTSNKGPRILKPCCKSTLTETKI EBSKGIKGGKV	2194
Db		1436	SSTIS-GSTCHITMSKASSGSTSPSNSQTSGSTVTMGSSST-----SGV-----	1478
Qy		2195	YKSLITGKVRNSEISGQMKOPIQANMPISRGRTWIHIPGVNNSSSSTSPVSKKGPPLK	2254
Db		1479	-----STGSASSTQPMQSTSOGSSAGSTV-----ASTAGPAA-----	1511
Qy		2255	TPASKSPSEQATTSPRAKBPVKSELPAVARQTSOIGGSKAPRSSGRDRSTPSRAPQ	2314
Db		1512	--SSTAPSSTGMTSSTSGTVGSTISESSTTASASQGTGTVTMGSSSTSGVSTSSASST	1569
Qy		2315	QPLSRPIQSPGRNISIPGRNGISPPNKLSQLPTTSPSTA-----STKS GSGSGMKVSTSP	2365
Db		1570	QP-----QWSTSQG-----SSAGSTVASSTAGLVSTSTVPSTSTMGTSTSS	1610
Qy		2370	GRQMSQNLTKOTGLSKNASSIPRESASKGLMNNGANKVELSRMWSMTKSCGES	2429
Db		1611	G-----TWGSTISESSTTASASQGT-STVTMGSSSTSGVSTSSASSTQPMQSTS	1659
Qy		2430	DRSERPVLVRQST----FIKEAPSPILRRKLKBESAFESLSPSRPASPTRSQATPVLS	2485
Db		1660	QGSAGSTVASSITGLVSTSTVPSTSTGMSTSSGTVGSTISESSTTAASASQGTGVTM	1719
Qy		2486	PSLPDMLSLTHSSHVVQAGWRKLPNLSPTTYENDGRPAKHDDIARSHSESPRLPINRS	2545
Db		1720	GSSTSGVST-SSASG-----QPMQSTSQGSAGSTVTVSSTASPAASTAPSTSG	1766
Qy		2546	TWKRHH-----SKHSSSLPRVSTRWTRTGGSSSSTLSASSESSEKAKSDEKHNVISGTHQ	2600
Db		1770	TMSSTSGTVGSTMQSSSTAASITSHGTVITLGSSSTSNQWSTSOGSSVGSIVASSTA	1829
Qy		2601	---SKENQVSAKGTWRKIKENEFSPTNSTSTQTVSSGATNGAESKTLIYQMAVAKXTEDV	2657
Db		1830	GLVSTSTVPSTSTGMTGSTSGTVGSTISESSTTASASSQGTGVTM-----	1875
Qy		2658	KVRIEDCPINNPRSGSPCTGNTPVDIVSEKANPNIKDSKDNCAKANGVSGVPMRTVG	2717
Db		1876	-----GSSSTSG-----VSTSSASSTQPMQSTSQGSAGSTVASS-----TAG	1913
Qy		2718	LENRLASFIOVDAPDOQKGTGIKFCQNNPVPVSET-NESSIVERTPFSSSSSKHSSPSGT	2766

Db 1914 LVSTST-----VPSSTGT---MGSTSGTVGTSISSSTAAGTSSTGSGTVTIGTSGT 1964
 QY 2777 VAA---RVTPNPNYSPKRSADSTSRPQIETPTVNNNTKKDSKTDST-----2823
 Db 1965 NPSPRLSLQITITPSPQSTESTQTSLPSSSPSTHSVSSSEGTMTSGATTSGDKMS 2024
 QY 2824 --ESSGTQSPKRRHSGSYLVTS 2842
 Db 2025 FLSSTGTITVFSRRGSLATT 2045

RESULT 8
 T23327
 adenomatous polyposis coli protein 1 - Caenorhabditis elegans
 N:Alternate names: apr-1 protein
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T23327; T03822
 R:Gardner, A.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19727
 A:Accession: T23327
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1186 <WIL>
 A:Cross-references: EMBL:Z19727; PIDN:CAB00045.1; GSPDB:GN000019; CBSP:K04G2.8a
 R:Experimental source: clone K04G2
 R:Rocheleau, C.E.; Downs, W.D.; Lin, R.; Wittmann, C.; Bei, Y.; Cha, Y.H.; Ali, M.; Prie
 Cell 90, 707-716, 1997
 A:Title: Wnt signaling and an APC related gene specify endoderm in early C. elegans embr
 A:Reference number: Z15051; MUID:97433081; PMID:9288750
 A:Accession: T03822
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1186 <RQC>
 A:Cross-references: EMBL:AF013950; NID:G2338717; PIDN:AAC47747.1; PID:G2338718
 A:Experimental source: strain Bristol
 C:Genetics:
 A:Gene: apr-1; CBSP:K04G2.8a
 A:Map position: 1
 A:Introns: 61/3; 113/2; 355/3; 551/3; 1067/3; 1161/3
 A:Note: apr-1

Query Match 3.5%; Score 511.5; DB 2; Length 1186;
 Best Local Similarity 22.0%; Pred. No. 1.8e-13;
 Matches 286; Conservative 217; Mismatches 580; Indels 219; Gaps 50;

QY 271 MATSGNGQSTTRMDHETASVLSSTHSAAPRLTSHLGTKVEMVYSLLSMLGTHDKDM 330
 Db 1 MSSSSDENETT--IHRGTSNTGGSGIYQPRAGSKRTSNVRHDVS-----DVDDE 50

QY 331 SRTLLAMSSQDSCSNRQSGCPLLIQLLHGNDKDSVLLGNRSGSKSKEARASAAALHN- 389
 Db 51 EEHARF--REDTAIEV--DDAITVLSLHFHKKEDIVPTDEDDNK-----LRELHEK 100

QY 390 ---IHSQPDKGRREIRVHLLEQITRAYCETCEWQEAHEPGMDQKNPAPVEHQI 446
 Db 101 IFALITSESVDNRKRLKALPASNCVR-----EQVYV--LRKKESTPPASYHEL 149

QY 447 CPACVLMKLSFBEHRHAMVELGGLOAIAELLQVDCMYGL--TNDHYSITLRRYAGMA 504
 Db 150 NAALHTIVKESFGEEYRKVATVGLVLEALAEVLEVHTFGINETNPGSHRNIRKLIANA 209

QY 505 LTNLTDFDANKATLCMKCMKALVAQLKSEDEQQVIASVLRNLSWRADVNSKTLR 564
 Db 210 LTNLTVQIHSKRLCSYDGFRCVV-RIVIESPNITQVYAGLRNLSWADSGSEALQ 268

QY 565 EVGSKALMECALEVKES--TLKSVLSALWNLSAHCTENKADICAVDGAALFVLTLYR 623
 Db 269 P--TVHALSIAVAHAHTRFDVTATLSALWNLGASHVENKRTICDTPNCLKVLASLLSPD 326

QY 624 SQNTLAIIESGGILRNVSLLIATNEDHQILRENNCLQTLHLKSHSLIIVSNACGT 683

Db 327 AFTSL--VDSATGILKVVSOVLANTSHTLEL--RSLITRMLTLKKSASFTCVNTLGA 382
 QY 684 LWNLSARNPKQDAL--WDMGAVSMLKXLIHSHKXKIAMGSAALRNLMANRPAYKANI 742
 Db 383 TANLTVKDPHQMIRQDMAAVQQLNVLNRNRRDITAVKSVLNTL--NOPCSHRYGDM 440
 QY 743 M-SGSSILPSLHVRRKQKALEAELDAQHLSSETFDNIDMLSPKASHRSQRHKQSLYGVYF 801
 Db 441 SHSVGGATGMQLSEPOLQMQ--TSHHAYHGTASPRLLSLRATRASPGKIYQQAQQLI 499
 QY 802 DTNRHDDNRSD---NFNTGNMTVLSPYLNITVLPSSSSSRGSLDSSRSEK--DRSLEREG 857
 Db 500 QTPQVDQRSSSLPRHF-----AVQRNGFVMAQSYNQ--MDHQOQQQMIYLOQOQQ 549
 QY 858 IGLGNYHEPATENPGTSSKRGQLQISTTAAQIAKVMEEVSAIHTSOEDRSSGTTTELHCV-- 915
 Db 550 I-----MTEDQAMEHH-----QQIMYLOQOQQQFHQIQOQQOMKQAEADPVPP 594
 QY 916 TERNALRRSSAHTHTNTYNTFKSENSNRTCSMPYAKLEYKSSNDSLNSVSSDGYGK 975
 Db 595 TDDLDIPTSTVMGTRGNSRSLGSMNPGSVMTWNSSLDTAANSSRALSPVSYND--IPA 653
 QY 976 RGQMKPSIESYSEDESKEFCSYGOYPADLAHKIH---SANHMDNDGELDTPIN----- 1026
 Db 654 SPTWCAQVFNLPKTESE---HHOLTQQQNTTHYSSGSANTMTSGATVPMDNITP 710
 QY 1027 -YSLK---YSDEQLNSGRQSPSONERWARPKHIEDEIKQSEQ-----RQSRNQS 1072
 Db 711 TVAILNPLVHEQTPNGTPRKTSSELDPSDDVLPGPSLEEEGDYALIGAAQKTDEL 770
 QY 1073 TTYPVYVE--STDDKHLKQPHFG-----QCEVSPYRSRGANGSETNRV--GSNHI 1121
 Db 771 LTRSIQSEMTSSTTPKMKVSPRLNGFFPTQKTTSPPAWSH-----PDTSPKSSSHRT 826

QY 1122 NQVYSQSLQCDDY-----EDDKPTNYSERSEEQHEEERPTNYIKYNEEKHFVD 1174
 Db 827 QPNRRQDASDADRLMESIMSEPKSRIISPRLAGTQOYLEPE--PERRSHSKNEEADRD 885

QY 1175 -----QPIDYS-LKYATDIPSSQKQSFSSKSSGSSQSKTEHMSSENSTSTPSSNAKR 1227
 Db 886 AFTASHEPDSHNGIDVARGSDWSPQQLHRMESLESQASSEDSFGLTAEPNSSTGSA-- 943

QY 1228 NQLHPSSAQSRSQPKAATCKVSSINQETIQTYCYVEDTPICFSRCSLSLSLSAED 1287
 Db 944 -----AANTMRFDDEIDASLPMDCVDDDDYD-----TYDHPEDYDEE 982

QY 1288 GCNTOQASANTLOIAIEKEKIGTSAEDPVEPAVQHPRTKSSRLQGSLSSES 1347
 Db 983 DDPDAT--QFDGVDQAULTIDCSMISSGSSGSRNETTTTTRDSKALATSTPKGSASSLP 1040

QY 1348 RHKAVERSSGAKS---PSKSG-----AQTEKSP-----SHYVQETPLMF 1384
 Db 1041 VRQATRVSTNGKSLPVKTINGSLVDKNPKPIIASRRPLRPKPTLLKDKHYPEE----- 1095

QY 1385 SRCTSVSLSDFSERS-----IASSVQSEPCSGMVGIIISPSDLPDP--GQTPPSP 1435
 Db 1096 -----DSIENQTRDDTIYVNAVVEAEQERIYVNAKQKNIQSPSISGNSPIAK 1146

QY 1436 SKTPPP-----PPQATQTKREVKNKAPTAKRESGPKQAAV 1472
 Db 1147 SAIVTPYNOKPPPTGRNNGEMNEKSVT-----PNPKOMLV 1183

RESULT 9

T23330
 hypothetical protein K04G2.8b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23330

R:Gardner, A.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z19727

A:Accession: T23330
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1188 <MIL>
A:Cross-references: EMBL:Z75712; PIDN: CAB00048.1; GSPDB: GN00019; CESP: K04G2.8b
A:Experimental source: clone K04G2
C:Genetics:
A:Gene: CESP:K04G2.8b
A:Map position: 1
A:Introns: 61/3; 113/2; 355/3; 551/3; 1069/3; 1163/3

Query Match 3.5%; Score 509.5; DB 2; Length 1188;
Best Local Similarity 22.0%; Pred. No. 2.2e-13;
Matches 286; Conservative 217; Mismatches 582; Indels 217; Gaps 50;

Qy 271 MATSGNGGSTRWDHTASVLSSSTHAPRLTSLHGTVMVYSLLSMLGTHDXXM 330
Dy 1 MSSSSDENETI--IHTGNTGSGGIYQPRAGSKRTSNVRHVS-----DVVDE 50
Qy 331 SRTLLAWSSQDCISFRQSCPLLIQLLHGNDKSVLLGNSRGSKEARARASAAALHN- 369
Dy 51 BEHYARF--REDALEIV--DDAITVLLSSLFHEKRDIVPTDEDDNK-----LRELHEK 100
Qy 390 ---IHSQDDKRRREIRVHLLEQTRAYCETWEQBEAHEPGMDQKMPAPVEHQI 446
Dy 101 IFALITESDVNRKRLKALPASNCVR-----EQVY--LRRKSPTPPASYHRL 149
Qy 447 CPACVLMKLGFBDEHRHANNELGLOAIAELLQVDCMYGL--TNDHYSITLRYAGMA 504
Dy 150 NAALHTIVKESFGEYRKVAIVLGLVEALAEVLILEVTFGNETNPGEHNRKLIANA 209
Qy 505 LTNJTFGVANAKTLCMKGCVRALVAOLKSESDLOQVIASVLRNLNWRADVNSKTLR 564
Dy 210 LTNLTGQIHSKRLCSYDGTIRCV--RVIESPNITQVYAGLRNLSWADSGMSALQ 268
Qy 565 EVGSKALMECALEVKKES--TLKSVLSALWNLASHCTENKADICAVDGAFLAVGTLYR 623
Dy 269 P--TVHALSIAVAHATHRFDTATLSALWNLASHVENKTICTDPCNLKVLASLSPD 326
Qy 624 SQTNTLAIIEGGGILRNVSLLIATNEDHQTLRENNCLTLOHLKSHSTIYVSNACGT 693
Dy 327 ARFTSL--VDSATGILKVVSYLANTSTHLEI--RSLLIIRMLTLKLSASFTCVNTLGA 382
Qy 684 LWNLSARPKEQBEAL--WDGAVSMLKNIHSHKHKMIANGSAALENLMANPFAKYKANI 742
Dy 383 IANLIVKDPHMQMIRQDMAAQQNLNLRNRRDDIRTAVKSVLNTL--NPFCSHRYGDM 440
Qy 743 M--SPGSSPLSHVRKOKALEAELDAQHLSFEDNIDNLSPKASHRSKORHKQSLYGYVF 801
Dy 441 SHSVGGGATGCMQLEPQLMQ--TSHAYHGTASPRLLSLRATRASPCKYIQPQAQQQLI 499
Qy 802 DTRHDDNRSD---NFNTGNMTVLSPYLNTTVLPSSSSSRGSLDSSRSEK--DRSLERBG 857
Dy 500 QTPQVDQRSSSLPRHF-----AVQRNGFYMAQSYNQ--MDHQHQQQMIYQLQQQQQ 549
Qy 858 IGLGNYHPATENPGTSSKRGQLISTAAQIAKVMEEVSAIHTSQDRSSGSTTELHCV-- 915
Dy 550 INF-----QTEQAQMEH-----QQIYVLOQQQQQFHOIQQQQQMAQEAQADVPFP 596
Qy 916 TDERNALRRSSAAHTSNYNTFKSENRTCSMPYAKLEYKRSNDSLSNVSSSDGVGK 975
Dy 597 TDDDLIDPTSTVWGTGRNSERSILGSMNPGSVMTNWNSSLDTAANSRALSFPVSYND-IPA 655
Qy 976 RQMKPSTIESYSEDESKFCGYPADLAHKH-----SANHMDNDGELDTPIN----- 1026
Dy 656 SPTMCAQVFNLPKSTSE---HHQTSQQNTTHYSSSGSANTMTSDGATTVEMDNIITP 712
Qy 1027 -YSLK---YSDQLNSGRQSPQNRWARPDKHIEDEIKQSEQ-----RQSRNQS 1072
Dy 713 TVAILNPILVHQETNGTVPRKTSBELDPSDVLPCPSLEEEEGDYAIIIGGAQKTDEL 772
Qy 1073 TYPVVTB---STDDKHLKQHFHG-----QOECVSPYRSGANGSETNRV--GSNFGI 1121

773 LTRSIQSEMPSTSSPTPKMKVSPRLNGFFSPTQKTTSPAMSH-----PDTSPFKSSSHRT 828
1122 NONVSQSLQCEDDY-----EDDKPTNYSBRYSEEOHEERPTNYSIKYNEEKRHVD 1174
829 QPNRRQDASDADRLMESIMSEMPKRIISPLAGTQVLEPE--PERRSHSKNEEADRRD 887
1175 -----QPIDYS--LKYATDIPSSQKQSFSSKSSGQSKTEHMSSESSSENTTPSSNAKR 1227
888 AFTASHEPDSHDGIDVARGSDWSPQQLHRMESLESQASSEDSPGLTAEFNPSTSGA-- 945
1228 QNQLHPSSAQSRSQPOKAACTCKVSSINQETITQYCVEDTPTICFSCSSLSLSAEDSI 1287
946 -----AANTWRFDEIDASLDMCVDVDDYD---TYDFEYDEDE 984
1288 GGNQTTQEADSAANTLQIAIEKIGTRSAEDVSEVPASVQHPRTSKSLQSSLSSESA 1347
985 DPDAT--QFDDGVDAGLTIDCSMISSGSSGSSORNETTTTSDRSKALATSTPKGSASSLP 1042
1348 RHKAVEFPSSGAKS-----PSKSG-----AQTPKSP-----EHYVQETPLMF 1384
1043 VRQATRVSTNGKSLRPVPTKNGSLVDXNPKPIIASRRPLPPLPKPTLLKDKHYPEE----- 1097
1385 SRCTSVSSLDSEFSRS-----IASSVQSEPCSGMVSGIISPSDLPDSP--QOTMPPSR 1435
1098 -----DSIENQTRDDTIYVNAFVVEAFQERIVMNAKQKQKIEQSPSINGSGPIAK 1148
1436 SKTPPP-----PPQATQTKREVPKNAKPTAEKRESGPKQAAV 1472
1149 SAIVTPYNTQKFPPTGRNNGEMSEKSVT-----PNPKQMLV 1185

RESULT 10
A43359
microtubule-associated protein MAP1A - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C:Accession: A43359; S22108
R:Langkopf, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
J. Biol. Chem. 267, 16561-16566, 1992
A:Title: Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messenger
A:Reference number: A43359; MUID:92355629; PMID:1379599
A:Accession: A43359
A:Molecule type: mRNA
A:Residues: 1-2774 <LAN>
A:Cross-references: GB:M83196; NID:G205837; PIDN:AAB48069.1; PID:G205538
A:Note: sequence extracted from NCBI backbone (NCBIN:111039, NCBI:P.111040)
R:Cravchik, A.
submitted to the EMBL Data Library, June 1992
A:Reference number: S22108
A:Accession: S22108
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 73-364, 'NLRSL', 370, 'QKN', 374, 'PSPKGL', 381-751, 'BSMMSQMAQR', 764, 'D', 766, 'LRI',
'WLKRNCPQPRQSP', 851, 'V', 853, 'NSL', 855, 'LPHRWLRIN', 865, 'W', 867, 'HSQLPDGGD', 877, 'Q', 879, 'I'
A:Cross-references: EMBL:X66840
A:Experimental source: strain Sprague Dawley
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein

Query Match 3.3%; Score 476; DB 2; Length 2774;
Best Local Similarity 18.6%; Pred. No. 1.6e-11;
Matches 540; Conservative 364; Mismatches 1038; Indels 954; Gaps 130;

Qy 609 VDGALAFVGLTGLVRSQT-----NTLAIIE--SGGGIL-----RVSSSLIATN 649
Dy 1 MDGVAEP-----SEVSETVDVPFDFLLEPTTSGGFLKLSKPCYIFPGGRGDSALFAVN 56
Qy 650 -----BDHRQILKENNCLOTLLQHLKSHSLTIVSNACGTLNLSARNPKDQEAQ 700
Dy 57 GFNILDVGGSDRK-----SCFWKLVRHL-----DRIDSVLLTH 89
Qy 701 MGAYSM--LKNLIHSHKMIANGSAAALNLMANRPAYKVD--ANIMSPGSSLSPLHVV-- 754

Db 90 IGADNLPINGLQK-----VALEBEQSGSSYSYDWMVKNLISPGLVVFVFNVPD 141
 QY 755 -----RKOKALEADAOHLSETFDNINLSPKASHRSQR--HKQSLYGYVYF 801
 Db 142 KLRLPDASRAKASIEBACLTQHNLRLGLOAEPLFRVSVNTIEPLTLPHMGVGRLDWY 201
 QY 802 DTRHNDNRSDNF-----TNGNMVLSPLYLNTTV-----LPSSSSSR 838
 Db 202 VLNPVKDSKEMQFLMQWAGNSAKTIGVLNKGAEISVPYLSITSLVWLPANPTEK 261
 QY 839 -----GSLDSSR-----SEKDSLE-----REGIGLG-----NYHATENPFTSSKR 876
 Db 262 IVRVLPFNAPQNLKLEGLKRLHLOFLRYPVATQDLAAGAVPANKPSKIKHRADSKE 321
 QY 877 GLOIS--TTAAQIAK--VMEEVSAIHTSOEDRSSGTTTELHCVTDERNALRRSSAAHTH 931
 Db 322 SLKAAPKTAVSKLAKREEVLEEGAKEARSELAKELAKTEKAKEPSEKPEKPSKSERV 381
 QY 932 SNTYNFTKSEN-----SNRTCSMPYAKLEYKRSS-----NDSLNSVSSSDGY-----973
 Db 382 GESSEALAKERLLIKOKAGKGLKEKISKLEEKOKKEKEIKERKELKTEEGRKEEKK 441
 QY 974 -----GKRGOMKPSIBSYDESKFCVGYQPADLAHKIHSANHMDNDGELDTPINYS 1028
 Db 442 DAKDEKKTKEVKKLSKEDLKPF-----TPVRKTLYKAKAPGRVYDKGPA 491
 QY 1029 LKYSDEOLNRSRQPSQNE-----RWAPKHIID--BIKOSORQRNOST 1073
 Db 492 ARGEKELSEPRTPPAQKGAAPAAVSGHRELALSPEDLTQDFELKREELGLLAEQRD 551
 QY 1074 T-----XPVYTESTDDKHLK-----FOP-----HFGQOECVPSVRSGANSETNR-- 1114
 Db 552 TGLGEPLPADATEQHPNSAIIQVTFPSGVLGEHVEREKEVVP--DSPGDKGS--TNRGP 609
 QY 1115 -----VGSNHGINQVNSQSLCOEDDYEDDKPTNYGERSSEEFQ 1152
 Db 610 DSGAEVEKEKETWEERKQREAEELGPENTAAREBEAEVKEVIEKAELEMEBETHPSDEE 669
 QY 1153 HEEERPTNYSIKYNEEKHVQPIDYSLKYATDI-----PSSQKQSFSSKSSSG 1203
 Db 670 GETKAESEFY-----QKHTQALKASPKSREALUGRDLGFGKAPKEKETASFUSSLAT 722
 QY 1204 QSSKTEHMSSESSSENTSTPSNKR--ONQLH-----PSSAQSRSGQ-----PQ 1244
 Db 723 PACATEHVSYIODETIPGYSETEQTISDEEIHDEPDERPAPPRFTSTVYLSGPGPGPP 782
 QY 1245 KAATCKVSSINQTIQTYCEDTPICFRCSLSSLSAEDIEIGNQITTOEADANTLOI 1304
 Db 783 EASQAADSAPASSSKTYGAPETELYP--PNMVAAPLAEER-----HVSSATSITE 832
 QY 1305 AETKEKIGTRSAED--PVSEVPASVQHPRTKSRLOGLSSLSSEARHKAVEF-----SSGA 1358
 Db 833 CDKLSLFSFASVADQSVASLTAPCTETCKSSLILDTVTISPSRTEATQGLDYVPSAGT 892
 QY 1359 KSPSKSGAQT--KSPPEHYVQETPLMFRCSYVSLSDFS-----SRSIASS-----1404
 Db 893 ISPTSSLEEDKGFKSP-----CDFSVTGESEKKGTVGRGLSGERKAVGKE 939
 QY 1405 -----VOSEPCSGMVSGIISPSDLPSPGOTMPPS-----RSKTPPP 1441
 Db 940 EKVVVTSEKLSGOYAAVFG-----APGHTLPPGPALGEVEERCLSPDDSTVWASPPP 993
 QY 1442 --PQTAQTKREVPKNKAPTAEKRE-----SGPKQA 1470
 Db 994 SGPPPSAAHT--PQHQSVPEDKSEPRDQEDSWGETKHSQGVSKEDGEEQVTKPGPEG 1049
 QY 1471 -----AVNAAVQ-----RVQVLPDAATLLHFATESPDPFGSCSSLSA 1508
 Db 1050 TSEEGKGPTRSPQADMPVSIAGGTGCTIQLLPEQDKAIVFTG-----EAGSNLGA 1103
 QY 1509 LSI-----DEFFIQDKVELR-----IMPPVQENDNGNETESEQP 1542
 Db 1104 GTLPGEVRTSTBEATEP--QKDEVLRFTDQSLSPEDAELSLSVSVSPDTTQKATPRSP 1161

QY 1543 KESNENOEKEAEKTIIDSEKDL---LDDSDDDDIIEILBECIISAMPTK-----SSRKAKK 1593
 Db 1162 CSLKEQO-----PHKDLWPMVSPEDTQSLSFSEE-----SPSKETSLDISSKQLSP 1207
 QY 1594 PAQTASKLPPPVARKPSQLPVYKL-----LPSQNRLOPQKHVSFTPGDMP-- 1639
 Db 1208 ESLGTLOPGLNLGKEERGPVWKAEDDSDCHLAPVSIPEPHRATVSPSTDETACTPAGGS 1367
 QY 1640 -----RVYC---VEGTPINPSTATSLDITIEPPNELAA-----GEGVGG 1678
 Db 1268 FSHGALSVDKXHSGEITPGGHFMT---SDSLTKSPESLSPPAMEDLAVWEGKAPG 1323
 QY 1679 AQSGEFKRDY-----IP-----TEGRSTDEAGGKTSTVTIPELDNKN 1717
 Db 1324 KEKEPELKSETROQKQILPEKVAVVEQDLI IHQDQALDEENKPGRQDKT--PEQKGRD 1382
 QY 1718 ABECDILAECINSAMPKSHKHPRVKIMDOVQOASASSAPNKNQLDGKKKPTSPVK 1777
 Db 1383 LDEXTAAELDKGPEPK-----EKOLDREDQGRAGPAPKAKASEQRTDLOQT- 1432
 QY 1778 PIPONTEYRVRKRVNADSKNNLNAERVSDNKKQNLKNNKDFNDKLPNNEDRVRS 1837
 Db 1433 ---QATEPRDAQERRD---SEED--KSLELADRTPEEKDRILVQEDR--- 1473
 QY 1838 FAFSPHHYPIEGTPYCFRNDSSLDFDDDDVDLSREKAEKRAKE---NKESEAKV 1894
 Db 1474 ---APEHSIP---EPTQTDRAFDKGTDDKEKEEASEEKEQVLEQKQWALKEGET-- 1524
 QY 1895 TSHTELSNQOSANKTQAIKAPINRGQPKPIQKQSTFPQSKDIPDRGAA---TDEK 1950
 Db 1525 -----LQOEARTAEQDETILKEKTKQCKSSFVEDKTT--TSKETVLVDQKSAKADSVQ 1577
 QY 1951 LQNAFIENTVCFSHNSLSLSLSDIDQENNN--KENEPFI---KETEPPDSQGP- ---S 2000
 Db 1578 QDGAALKETKALGLEESPAGSKAREQEKYKWEQDVVQGWRETSP--TRGEPVGGQKEP 1635
 QY 2001 KPQASGVAPXS--PHVEDTVPVCFRNSLSSLSLSDIDEDDLQECI--SSAMPK--KKKPSR 2055
 Db 1636 VPWEGKSPQEVYWRDRDITLQDAYWRELCDKRVWPPHELDGQGARPRYCEEREST 1695
 QY 2056 L--KGDNE-----KHSPRNMGILGEDTLDLKDQRPDSEHGLSPDSENFQWKAIOEG 2107
 Db 1696 FLDEGPEDEITPLQHTFRS-----PWTSDFKDFQEPULQKGLVER---WLA--- 1740
 QY 2108 ANSIVSSLHQAACLSRQASDSDSILSKSGISLGSFPFLTPQEEK---PF--- 2159
 Db 1741 -----ESPVGLPPEEDKLTSPFEIIS 1763
 QY 2160 -----TSNKGPRILKPGEKS---TLETKKIESRK----- 2186
 Db 1764 PPASPPMTQORVPS--APQESPVPTTESTAPWRNEPTTPSWLAETPPWPKDRPLPPA 1821
 QY 2187 -----GKGGKXVYKSLITKGVRSNSEISGOMKQPLQANPISIRG 2227
 Db 1822 PLSFAPAPPTPAPEPHTPVPSWGLAEDYSVAVQEAEGGYPSPGLKDYKRAEGE 1881
 QY 2228 RTMHIHPGVNSSSSTSPVSKGPPPLKT-----PASKSPSE----- 2263
 Db 1882 REBEGGAGADSSFSFKVPEAGESLATRDTQTEPEQREPTPYPPERSFOVADIEQOM 1941
 QY 2264 ---GQIATT--SPRGAK---PSVKSELSFVARQTSQIGGSKAPSRSGSRDSTP--- 2309
 Db 1942 LTGLGPACPTREPPPLGASGDWPPHLSKTEBAAGCNTSAEKETSSPASQNLQSDTAPFSY 2001
 QY 2310 ---SRPAQQLSPIOSQGRN---STSPG-----RNGISPPNKLSQLPRTSSPT 2353
 Db 2002 ASLAGAVPPRQBP--DPGNVTFPSITPPAVPPAPISLSKULSPPLNGSTVSCSPDRIT 2059
 QY 2354 ASTKSGSGRMVSTSPGRQMSQOQLTKTGLSKNASSIPRSESASKGLNQ---MNGNGA 2410
 Db 2060 PSPKETGRGHW-----DDGTNDSLEKAGREQEKE"RSPSPHPMPMGHSSILW 2108

Qy	2411	NKVELSRMSTKSSGS-----	ESP	2430
Db	2109	PETAYSLSDSHLGSVRSLDPPAGAFGSSLQAPPOLPAPRSPAPCSLAFSGD		2168
Qy	2431	RS-----ERPVLVRQSTFIKEAPSTLRRKLESASFES-----LSPSSRPASP-----	T	2475
Db	2169	RALALVPGTPTRTDHEVLEVTAKPSLDSSLPOLPSPSSPGCLLNLPRAPSPALSEGS		2228
Qy	2476	RSQQTPLVPSL-----PDMSLTHSSVQAG-----WRKLPNLSPITIEYNDGR		2522
Db	2229	SSEATTIVISVABRFPFGLAEAEQSAEGLSGKESAAHSLDLTP--LSPA-----	P	2279
Qy	2523	AKRDIARSHSESPRLPINRSGTWKREHKGHSSLP-----RVSTWRRTGSSS		2571
Db	2280	SASLDLAPAPAPAPAP-----GLPDLGDGTLPCRPECTGELTKPS		2323
Qy	2572	SILSASSESEKAKSEDEKHVNS---LSGYQSKENQV---AXGTWRKIKENEFSTN-		2624
Db	2324	PFLSPSGD--HEANGPGETSLNPPGFVTATAKEEAEAPHAWERGSWPEGAERSSRDTL		2381
Qy	2625	-STQSTVSSGATNGAESKTLIYQ-----APAVSKTEDVWV		2659
Db	2382	LSSQPLRPGKSSGPPCSLSEVEAPQGCATDPRPHCGELSPFLNPLPSTDDSDL		2441
Qy	2660	RIEDCPI-----NNPRSGR-SPTGNTPVIDSVSEKAMPNIKDKNDKQAKQNVGNSVPMR		2714
Db	2442	STEARLAGKGRRRVGRPGATGCPMAD---ETPPTSASDSSGSSQSDSDVPPEBEC		2498
Qy	2715	TVGLENLNS-----FIQVD-APQKGT-ETKPGONNV-----PV		2748
Db	2499	SITAEALDSDEGDFLVDKAGVSGTHHPRGHDPPTPLDPRSPRPPDVCMADPE		2558
Qy	2749	SETNESSIVERTPFS-----SSSSKHSSPSGTVAAARVTFPNPSPRKSADSTS---		2799
Db	2559	GLSSSGRVERLREKGRPRAPGRAPKAPASPARLDIR---GKRSPGKGPVDRTSRTV		2615
Qy	2800	ARPSQIPTPVNNYK 2815		
Db	2616	PRPRSTPSQVTSABEK 2631		

RESULT 11

hypotheical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34513

R:Favella, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Accession: T34513

A:Molecule type: DNA

A:Residues: 1-3507 <FAV>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match

Best Local Similarity 3.1%; Score 454; DB 2; Length 3507;

Matches 465; Conservative 277; Mismatches 849; Indels 740; Gaps 100;

Qy 1051 RPKHTEIDIKSEORQRNQST-----TYPVYTESTDDKHLKFOHFQGCQCV 1099

Db 629 RSNHSTSDLIQTQVQQRNFSTGQILLTRGVSSGEAVTQTDADEF----- 676

Qy 1100 SPYRSRANGSETNRVGSNHNQNVQSLOCEDYEDDKPTNYSERVSEEQHEE----- 1155

Db	677	-----GLEISAADLAGSGSGITLPTT-----LEPKIEGSKKASGVWTEDEGEDELM		726
Qy	1156	BERPTNYSIKYN-----EKRHVQDP-----		1176
Db	727	EGSGSWSTTLNGTIGITSPRSEGTIRVIRITLGEDGEPETATKPGISAPDKTGEKSXTE		786
Qy	1177	IDYSLKYATDIPSSQKQSFPSKSSGSSQSKTEHMSWSSSENTSTPSSNAKRONQLHPSSA		1236
Db	787	SDGEKELTVEKDGEAQSGSGSATSGKSKSEATSGSSSSAKSGTGEASGSSGASSSSG		846
Qy	1237	QSRSGOPKA-----ATCKVSSINQETICTYCVEDTPICFSCSSLSLSLSAEDTIG		1288
Db	847	SGVSGESGSSVTESGSGFTSGSGSVGSEATGTVGDSE-----SGKPSKSSSTEKLIP		900
Qy	1289	CNQTTQEAADANTLOTAIEIKKIGTRSAEDPVSYPVAVSQHPRTKSRLOQSSLSSESAR		1348
Db	901	FTKNGEKSPIGS-----DITGKESSEETTSRKPIEGSDSLTEGSG--SGEWFETGSG		952
Qy	1349	HKAVFESSGA-----KSPSKSGAQ-----TPKSP--PEHYVQETPLMFSRCTSVS-		1391
Db	953	H-----FESGKSVTSVTKGTQSGAEGSGSGKVPKPGAPE---ITDGESSSTSGD		1005
Qy	1392	-----SLDSFESRSIAS-----SVOS		1407
Db	1006	KSGKPADKSDKNVKTGDKNPDITTDGEDTSETSGGEGQPKGSKGQPPDGKSEVK		1065
Qy	1408	EPCS-----GMVSGI-----ISPDLDLP-----SPG-----QTMPPSR		1435
Db	1066	KPTSEVDGPNLSGTGKGNVPLKPTDLPESGILTSSGKNSTFEHGTKLRLPP--		1123
Qy	1436	SKTPPPPTQATKRVKPKNAPTAP---KRSGPKQAANAQVQVLPADLTLLHFA		1492
Db	1124	-KTEDKSSETPQGLGISAKKPEPDGTSKEVG-----LELMESTTSGSTT		1170
Qy	1493	TESTPDGFCSSLSALSLEPPIQ-----KQVELRIMPVQVQNDGNNE-		1536
Db	1171	LDSVSGLEISGDLTKATKPHVEIEGSGTGDEETATTRDVYSKTKKPRVEVDGNDG		1230
Qy	1537	-----TESEQPKSENEQKAEKTIQSEK-----LLDSDDDDDIDILEECII		1580
Db	1231	ETSGVDKPTTAPTPSSSAESTSRIPITSESPGSGEGAGVPSPGSGESSTAPD		1290
Qy	1581	SAMPTKSRKAKPAQATASKLPP-----PVARKPSQLPVYKLPSONRLQPKHVSFT		1633
Db	1291	GVFTSSATAPEVPTTSASSTPDVAEESGIPSSKTAEPLETTAPSTEVTSPEG--SGT		1348
Qy	1634	PGDMPRVYCVGT-PINFSTATSLDLTIESPNELAAEGVR-----GGAQSG		1682
Db	1349	EESTLPP--TEGSGESTTSSAPTVEFATVLPQNRNEKPEPTKDTFALPTTTTGAQAN		1405
Qy	1683	EPEKRDITPEGRSTDEA-----QGG--KTSVVTIPE-----		1712
Db	1406	DSSVENTKCT---SSDECGLDALCERRTGVCRCPEGAPPKSCVDVDECATGDHNC		1462
Qy	1713	-----LDDNKAEE-----GDILAECI		1728
Db	1463	ESARQNVGVGYACFTGFRKADDDGSCQIDICTEHNSTCCGANAKCVNKPPTYSCCE		1522
Qy	1729	NSAMPKSKHKPFVRKIMDVQOASASSAPKN-QLD-----GKKKKTSP		1775
Db	1523	NGFLGDGYQCVPTTKAPCDSTQSSKSHCESNMSCEVDVDSVECKECMGYKKSQV		1581
Qy	1776	VKPIQNTYRTRVRKNADSKNINLAERVSNDKSKQNLKNS---KDFND-----X		1826
Db	1582	CEDINECVAEKAPCSLNA---NCVNMNGTFS---CSCQGYRGDGMCTDINECDEHPC		1635
Qy	1827	LPNNE-DRVRGSPAFDPSPHY-----TPTEGT-----PYCFS---RND		1860
Db	1636	HPHAECTNLEGSFKCECHSGFGIKKCTNPJERSCEDVEKFCGRVDHVSLSVRYNG		1695
Qy	1861	SLSSL-----DFDDDDVDLSREKAELEKA-----KENKSEA		1892
Db	1696	SLSVCECEPCFRFEKBSNCSVDIDICEESRNNCDPASAVCVNTEGSRCECAEGEG		1755

Matches 382; Conservative 251; Mismatches 744; Indels 533; Gaps 78;
 QY 1210 HMISSSENIS-----TPSSNAKQNLHPSSAQSRSQPOKAATCKVSSI 1254
 Db 1 HSGSSSDTKVKPTPPRQSHSGSISYPKVKAQTPPGPSLSGSKPCP-----49
 QY 1255 NQETIQYCVEDTPICFRCSL--SSUSABDEIGN-----QTQEAASANTLQIAEI 1307
 Db 50 -QEKSDSLVQSCFSGSLCAGVKSTPPGESYFGVSSLLQKQSQSTSPDHRSDTSPEV 108
 QY 1308 KEGIGTSAEDPVEVPAVSHPET--KSSRLQSSLSSESARHKAVERFSSGAKSPS---1362
 Db 109 RQS-----HSEFSLQSKQTSFKGGRSRSSSPVTELSRSPQRDRGFASPM 159
 QY 1363 KSGAOTPKSPPEHYVQBTPLMFGRCTSVSLDFEGRSIASSVQSBPCSGMWGIIISP-S 1421
 Db 160 KSG-----MSPEQSRFQ-----SDSSYPTVDSNLSLQGRLETAESKEMALPQPE 206
 QY 1422 DLDPSPGQTPPPRSKTPPPPPQT-----AQTKEVPKNAKPTAEKRESGPKQAAVN 1473
 Db 207 DATASP-----PRQDKFSPFPVQDRPSSLVKDTLTPPRERSGAGSGSPETKQNSALP 262
 QY 1474 AAVQRVQL-----PDADTLHFAT-----ESTPDGFCSSSL-----ALSDEPFIQ 1517
 Db 263 TSSQDEELMEWKESEEPAGQILSHLSSELKENSTSNFESSPEVERPAVSLTLDQSQSO 322
 QY 1518 KDVELRMP-----PVQENDNGN-----ETESQPKESNENQKEAB 1554
 Db 323 ASLEAVEVPMASWGGPHFSPBEKLSNPLRENSFGSPLEFRNSGPGTGEMNTGFSSE 382
 QY 1555 KTDISEKDLDDSDDDDIETILECIIISAMPTKSRKAKPAQATKLPVPAVKRPSQLPV 1614
 Db 383 VKEDLNGPFLNQLTDPDLNKE-----QSTRS-----GHSSSELSDAVEKAG-----427
 QY 1615 KYLLPSQNRLOPKHVSFTDGDMPRYCVGEGTPIINFSTATS--LSDLTIESPPNLAAG 1672
 Db 428 -----MSSNQSISSPVLDAVPTPSRERS-----SSASPENMDGLPRTPPSRSSRG 474
 QY 1673 E--GVRGAQSGFEKRDITPTGRTDDEAQGKTSVTIPELDNKAEGDILACINS 1730
 Db 475 SFGLDGGS-----GTSRHSLSGS-----PGMKDIPRTPSRGRGECDS 515
 QY 1731 AMPKGSHKFRVKINDVQQAASASSAPNKQ--LDGKKKKPTSPVKPIQNTYEYTR 1788
 Db 516 PEPKALPQT-----RPRSRSFSPPELNKCLTPQERSGSE--SSVDQKTVAETP 564
 QY 1789 V--RKNADSKNLAERVEDNDKSKQNLKNSKDFNDKLNEDRVGSAFSPHY 1846
 Db 565 LGQSRSGSQELDVKPSAPQERSE-----SDSSPDSKAKTR-----602
 QY 1847 TPIEGTFCFSRNDLSLSDFDVDDVLDGREKAEKAKENKESAKVTSHTELTSNQOS 1906
 Db 603 TPLR-----QRSRSGSSPEVD-----SKRSLSPRESRSGSSPEVKDKPRAAPRQS 648
 QY 1907 ANKQAIKQPINRGOKPILQKSTFPQSSK--DIPRGATDEKLNFAIENTPVCF 1963
 Db 649 GDSDSPEPKAPAPRALP-----RRSRSGSSKGRGSPGEGSSTES-----SPEHPP---695
 QY 1964 SHNSSLSSLSDIQENNNKENEP--IKETFPDQSGEPKQASGVAPKSFHYED-----2016
 Db 696 -----KSTARRGSRSPKTKSTRTPPRRSRSSPELTRKARLSRRSRSSASSPE 747
 QY 2017 -----TPVCFSRNSSLSDIDEDDLQECISAMPKKKKPRLKGDNEKHSRPNMGIL 2072
 Db 748 TRSRTPRRHRSFVSPEPAEKSRSSRRRRGASSPRTKTSR--RGRSPSPKPRGLQSR 806
 QY 2073 GEDLTLDLQIDOPDSEHGLSPDSENFDMKATQEGANSIVSLHQAAAAACLSRQASDS 2132
 Db 807 SRSRRKKTITRRD-----RSGSSQSTR-----RRGRSR 839
 QY 2133 DSIILSKSGISGSPFH--LTPQEEKFTPSNKGPRILKPGKSTLETKKIBESKIGG 2191
 Db 840 SRVTRRRRG--GSGYHSRSPARQSSRTSSRRR-----GRSRTPTTSKRSRSTSPAP 892

QY 2192 KKVYKSLITKVRNSEISQMKOPLOANMPSISGRRTMIHPCGVRNSSSSTSPVSKGP 2251
 Db 893 WK-----RERSRASPATRRSRRTPLISR-----RRSRSTSPVSRRS 932
 QY 2252 PLKTPAKSPBEGQTATTPRCAPKSPKSELSPVARTOSTQIGSSKAP--SRSGSRDSTP- 2309
 Db 933 RSRTSVTRRRSRASPVSRRS--RSRTPPVTRRRSR-----SRTPTRRRSRSTPP 984
 QY 2310 -----SRPAQPLSRPIQPCGNSISP-----GRNGISPPNKLQPLRTSPSPSTASYSS 2359
 Db 985 VTRRRSRRTPPVTR--RRSRSTPTTRRRSRSTSPVTRRRSRRTSPVTRRRSR 1041
 QY 2360 GSGKMSYTPSORQMSQQNLT-----KOTGLSKWASSIPRSESASKGLNQNNGA 2410
 Db 1042 -----TSVTRRRSRRTPPAIRRRSRRTPLPKRGRSRSPAIRRRSRRTPTA 1094
 QY 2411 NKVELSR-----MSSTKSSGSESDRSR-----PVIV--RQSTFIKEAP 2448
 Db 1095 RKGSLTRSPAIRRRSAGSSSDRSRSATPPATRNHSGRTPPVALNSGRMSCFSRPSM 1154
 QY 2449 SPTLRRKLEESAFESLSPSRSPASPTRSQATPVL---SPSLPDMGLST---HSSVQA 2501
 Db 1155 SPTPLDRCSRCPMLEPLGSSRTPM--VLQAGGSMMDGPGPRIPDHQRTSVPENHQSRI 1213
 QY 2502 G-----GWRKLPNLSPTIENDGRPAKHDIARSHESPERLPIKNSGTWKRH-- 2551
 Db 1214 ALALTAISLGTARPPMSAA-----GLAARMQVPAPVPLMSLRTAPAAALA 1261
 QY 2552 -----SKHSSSLPRVSTW-----RRTGSSSILSASSSESEKAKSEKHYN-- 2593
 Db 1262 SRIPAASAAANLASARTPAIPTAVNLADSRTPAAAANLASPRTAVAFSA---VWLA 1317
 QY 2594 -----SIGTKOSKENQVSAKGT--WRKIKENFSP--TN 2624
 Db 1318 DPRTPTAPAVNLAGARTPAALALSLTSGSTPTTAAKYPSSSRTPQAPASANLVGPSAH 1377
 QY 2625 STQTSVSGATNG--AESKTLIYQMAPAVKIEDVWVRI-----2661
 Db 1378 ATAPNLAGRTAAALPASLTSAAPALSGANLTSFVPLSAYERVSGRTSPPLDRA 1437
 QY 2662 -----EDCPINPRSGRSTGN-----TPVIDSVSEKANPNKD 2696
 Db 1438 RSRTPPSAPQSRTSERAPSPSSRMGQAPSQSLPPAQDQPRSPVPSAFSDQSRCLIAQ 1497
 QY 2697 SKDQAKQNVGNGSVPRITVGLNENLNSFIOVDAPDKGTETI-----KPCQNNP-- 2745
 Db 1498 TTPVAGSQSLSSGAVATTTSAGDH--NGMLSVPAPGVPHSDVGEPPASTGAQOQPSALAAL 1556
 QY 2746 VPVBETNESSIVERTPPSSSSSSSKHSSPSGT-----VAARVTPPENYN 2787
 Db 1557 QPAKERSS 1613
 QY 2788 PSRKSSADSTSRAPSOIPTPVNNNTKRSKTDSTDESSGCTQSPKHHGS 2837
 Db 1614 PTP---APKEAVREGRPPEFTPAKRRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1659

RESULT 14

T30249
 cell proliferation antigen Ki-67 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30249
 R:Starborg, M.; Gell, K.; Brundell, E.; Hoog, C.
 J. Cell Sci. 109, 143-153, 1996
 A:Title: The murine Ki-67, cell proliferation antigen accumulates in the nucleolar and het
 1 for cell cycle progression.
 A:Reference number: 220787; MUID:96431717; PMID:8834799
 A:Accession: T30249
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-2938 <STA>

A;Cross-references: EMBL:X82786; NID:g1177527; PIDN:CAA58026.1; PID:g1177528
A;Experimental source: strain CBA; testis
C;Genetics: A;Gene: Ki-67
C;Keywords: cell cycle control; nucleus; tandem repeat

Query Match 2.9%; Score 420; DB 2; Length 2938;
Best Local Similarity 18.6%; Pred No. 3.3e-09;
Matches 604; Conservative 439; Mismatches 1268; Indels 932; Gaps 147;

QY 80 SNFPVKLRKXMSLSYSGREGSVSSRGECSPVMGSGF--PRRGFV-----124
DB 109 TFFPGKSLGKEPFRRA--SRDSFCADPDGEGDTKASKWTASRRSFVYAKGLSADSPASD 166
QY 125 ---NGSRESGYLEE-----LEKSRILLADLKEEKEKD 156
DB 167 GSKNSVSDSGSHVEQHTGRNIVETSGSLRLSPGLOCAVGTGRSLLPOTLOSNSNEKE 226
QY 157 WYIAQL-QNLTKRID-----SLPTENFSLQTDNTRRLQLEYEAR-----QIRVA 199
DB 227 SPFEKLYQSMKEELDVKSQKCRKSEPOPDRAABESRETQLLVSGRAAKSGGTPVTAA 286
QY 200 MEEQLG-----TCQMEKRAQR--IARIQIEKDIILRIROLLOSQATEAERS 245
DB 287 SSPKVGKIWRWRGMPVPTSTETAKMKTFRHSQQLKDEDSRV-----TCRRH 337
QY 246 SONKHETGSHDAERQEGQGEINNAV-----SGNGOGSTTRMDHETASVLSSSSTH 298
DB 338 SVNLDGSGAQAHVHTVTPG---KLATFNQTPVEAGVDGSPAUTPEH-----SSSPQR 387
QY 299 SAPRRL---TSHLGKVMYVSLSMGLTHDKDMSRTLLAMSSQSCISMROGCLPL 355
DB 388 SIPAKVEAPSAETQNLRLTLQRLVFGKKTPKGSFSPKEKLATAAEQTC-----SG-LPG 441
QY 356 LIQLLGHNDKDSVLLGNRSGKEARARASAAHL-----NIHQSOPDDKRGRRERLVL 407
DB 442 LSSVDISNFGSI--NKSEGMKRRRVSGHGLRPELFDENLPNTV-LRGTPTPKR 498
QY 408 HLEQIRAYCTEWQBAHPGMDQDNMPAPVHQICPAVCVLMKLSFDEHRHAMN 467
DB 499 SLGTHSPAVLTIK-ERPQSPGQKESPGITPRTN-----DQRRRGRT 542
QY 468 ELGGLOAELLOVDCEMYGJTNHYSITLRYAGMALNTLTFDGVANKATLC-SMGCM 526
DB 543 SSG-----SNLCETDIPKAGKSGNLP-----AKRASISRSQHGL 580
QY 527 RALVAQLKSESDLOQVIASVLRNLNLSWPAVY---NSKKTLEVGSVKALMECALEVKES 583
DB 581 QMICKRRSGASEANLIVAK-----SW-ADVVLGVKQTQTKVAKHVPPKQTSKRQRPS 634
QY 584 TLKSVLSALWN-----LSAHC-----ENKADICAVDG 611
DB 635 TPKKPTSLNHNQFTTGHSNPSCTIVVGRAQIEKVSVPARPYKMLNMLNLRKVDSEDS 694
QY 612 ALAFVLGLTVRSQNTLAIETESGGILRNYSLL-----IATNED-----651
DB 695 GLTEFKTPVKEKQQ-----MSDTGVSLSNANLSERQLOVNSGDPIPEPITILGEKV 750
QY 652 -----HRQILRENCLQTLQHLKSHSLTVSNACGTLNLSAR 690
DB 751 LSSTRNAAKQSDRYSGASFTLRRRSIKHENTVQT---PKVHNIT-----DLEKK 797
QY 691 NP-KDOEALWDMGAVMLKNLIHSKHKM-----AMGSAALRLMA-----731
DB 798 TPVSETEPLKTVASSVKLRRSRELRHLTVETVNEKTEAVLANTTARHLRGTTFREKQVDQ 857
QY 732 -----NRPKYKDIANIMSPGSLPSLH---VRQKALEAELDAQLHSETPDNIDNLSPK 782
DB 858 QVQDNENAPORCKESGELSEGSEKTSARRSSARKQKPTKDLGSGQWVTQTADYAEALLSQ 917
QY 783 ASHRKQKHQSLGYDVFDRNHDNDNSDFNFGNMTVLSPLYLNTVLPSSSSSRGSGLD 842
DB 918 G-----QGTIQNL-----ESMMQNTSISGEOG---ITEKKVNIIVYATKEXSPKTP 963

QY 843 SRSE-----KORSLERER-GIGLGNVHPATENPGTSSKRQLQISTTA 884
DB 964 GKQAQLEGFAGLKEHFETPNPKDKBITEDRTRVLCKSPQVTTTITTKP--QTSTG 1021
QY 885 AQIAKVMEEVSA---IHTSQEDRSSGSTTELHCVTDERNALRRS-----SAATHSNT 934
DB 1022 KKV-DMKESASALTKEIHPGESRHPKILKEC--EDIKALKQSENEMLTSTVNGSKRT 1078
QY 935 YNFKSEN---SNRTC---SMP---YAKLEYKRSNDSLSNSVSSDGYGKRGOMKP 981
DB 1079 LGSKKKAQPLEDLTCQELFISFVPTNIKKIPSPHTQPVRTASTKRLSKTGLSKV 1138
QY 982 SIESYSEDDSKCSYGOVPADLAHKHS-----ANHMDDND--GELDTPIVSLKY 1031
DB 1139 DVR-----QEPSTLCKTKSPGAPGTPAPVQEBNCTATWETP-----1177
QY 1032 SDQLNLSGRQSPQNERWARPKHI--IEDEIKOSEQORQSRNQSTTYPVYTESTDDXHLKF 1089
DB 1178 -KQLESIENTLGLRKQSRTPKXDTITGQDSFOIPDEHANG-----PLVVVKT--KKMF 1227
QY 1090 QPHFGQCECVSPYRSGANGSETRNVGSHNGINQVNSQLCOEDDYEDDKPTNYSERYSE 1149
DB 1228 --NSPOPESAITKSR-----EROSRASISKIDVKE-----ELLES 1261
QY 1150 EQHHEEERPTNYSIKYNEBEKRVHDQIDYSLKYATDIPSSQKQSFQSKSSSGQSCKTE 1209
DB 1262 EEHLQLEGVDITFQVSTNKVIRSRKPAKRLDSTAGMPSKEMRC---SSKDNTPCLE 1317
QY 1210 HMSSESENTSTP-----SSNAKQNLHPSSAQSRSQGPQKAATCKVSSINQET 1258
DB 1318 DLNGFOELFQMPGYANDSLTTGISTMLARSPQLGPVRTQ-----INRKS 1361
QY 1259 IQYCYVEDTPICFSCSSLSISSAEDGICNOTQOADSANTLQIAEI-KEKI-----G 1312
DB 1362 LPKILIRKMDV---TEEISGL--WKQSLGRVHTTQEQEDNAIKAMEIPKETLQTAADG 1415
QY 1313 TRSAEDPVSEVPVAVSQHPRTKSRLOGSSLSSESARHKAVERFSSGAKSPKSGAQTPKSP 1372
DB 1416 TRLTROP-----QTPKEKQVPLEDHSHVFOE--LFQTSRYCSDELIGNKQTRMSLRSP 1465
QY 1373 PEHVQETPLMFRCSTSVSLSDFESRSIASSVQSEPC--SGMVSGI-ISPDS-----LP 1424
DB 1466 QPFVFR-TPTSRLAK--TSVGNIAVREKISPVSPQCAGTEVVHIPIGEDDTEKNGVK 1523
QY 1425 DSPQGMPPSRSKTPPPPPQTAQTKREVPKNAKPTAEKRE-----SGPKQAQVNAVOR 1478
DB 1524 ESTPQILDSSASRTVSKRQQAHEER--PQFSGDLFHPQELFOTPASGKDPVTVDSTTKI 1581
QY 1479 VQVLPADTLLHPATESTPDGFCSSSSLSALSDEFFIOKDVLR-----IMPPVOENDN 1533
DB 1582 ALQSPQPGHIINPASVKRQSNMSLRKDMREFSI---LEKQTSQSGRDAGTPAPMOE-EN 1636
QY 1534 GNETESEOPKE-----SNEOKEAEKTIIDSEKDLDDSD-----DDDIETILE 1576
DB 1637 GTTAIMETPKQLDFFGNSTGHKRRPRTPKNRAQPLEDLGQELFOTPAGADPVSVEE 1596
QY 1577 ECTIS-----AMPTKSRKAKPAQTA-SKLPPVARKPSQLPVYKLLPSQNLQPKHV 1630
DB 1697 SAKISLASSOAEVPTPASTKRRSKTGLSKV--DVRQEPSTL-----1736
QY 1631 SFTPGDMPRVYVEGTPI-----NPFSTATSLDLTIBSPNLEAAGEVGRGQAQGEPE 1685
DB 1737 ---GRKMSLGRAPGTAPVQEBNDSA---FMETPKQKLD---TGNSSGHKR 1781
QY 1686 KRDTIPEGRSTDEAQ-----GKTSSTVIPE-----LDDNKAERGDILAEICNSA 1731
DB 1782 RPQTPKIRAQPLEDLGQELFOTPAGANDSVTVESVMSLESSQAE--PVKTPASTKR 1839
QY 1732 MPK-GKSHRPFRVKXIMDQVQOASASSAP-----NKMQLD-----GKKKK 1771
DB 1840 LSKTGLSKVDVREDPSILEKTKSPGTAPVQEBNCTAFMETPKQKLDFTGNSSGHKR 1899

1772 PTSP---VKPIQNTYRTVRKADSKNNLAERVFSDNKKQNLKN----- 1819
1900 PRPKIRAQPLEDLGFOELFQTPAGASDSVTE-----ESAKMSLESSQAKPVKTPA 1952
1820 -----SKDFNDKLPNNEDVRGSPAFDSPHHYTPTEGTPY-CFSRNDLSLSLDFDDDDVD 1873
1953 STKRLSKTGLSKVDVEDPSTLGKTKSPOR---AFGTPAPVOEENDSTAFMETKPKQLD 2009
1874 LSREKAELEKAKENKESEAKVTSHTELTGNQ--QSANKTQAIKQPINRGQPKPILQKOS 1931
2010 FA-----EN-SGSKRRSRTSKRSQPLEDLGFOELFQTPAGASNPVSVEESAK 2058
1932 TFOSSKIDIPDRGAATEKLFONFAIENTPVCFSHNSSLSDIDQ-----ENNK 1982
2059 ISLESSQAEPVTRASTKLSKTLGKNKDVREGH-SPLSKSCASQVMQTTILGEDHGR 2117
1983 EN-----EPI-----KTEPPDSQGPSPKQORSCKRSQSPEDLSGVQEVFQTSQHNKDSVTVDN 2177
2118 ETQDGVKLLAQLKLEPAIYVTRGKQORSCKRSQSPEDLSGVQEVFQTSQHNKDSVTVDN 2177
2017 TPVCFNRNSSLSDSIDSEDLLOECISSAMPKKKPSRLKGNKHSRNNNGGILGEDL 2076
2178 L-----AKLPSSSPPLEPTDTSVTSRRQARTGLRKVHVKN--ELSGGIMHPQI 2223
2077 TLDLKOIQRDSEHGLSPDSNFDMKAIQEGANSIVSSLHQAAAACLSQASSSDSITL 2136
2224 SGSEIVDLPR-----EPGEG---KVITRKQSVIRKLDTEVNVPRSKQRIKRAEKL 2273
2137 S-----LKSGLISLSPHLLTPDOEEKPF--TSNKGPRILK----- 2169
2274 EDLPFGFQELCOAPSLVMSVIVEKTP--KMPDKSPPEPVDTSQARRRLRLVWTEPI 2331
2170 PGKST---LETKKISESGIKGGKVKYKSLITKVRNSSEISGOMKOP--LQANMPSI 2224
2332 PQRKTRVVRQTRNTOKEPISDNQGMEEFKESSVQKQDPSVSLTGRNQPRIVKEKTQPL 2391
2225 SRGRTMIHFGVRNSSSTSPVSK-----KGPFLKTPASKSPS-BG 2264
2392 BELTSFOEETAKRISKSPOEKEKTLGLKQLRIQLINDGVKEP--TAQKQPSRET 2449
2265 QTATTSRPG-----AKPSVKSELSPVARQTSQICGSKKAPRSRSDSTSPRAQ--- 2314
2450 RNTLKPEVGDINVEEVKSKTKQIDPVASVP-----VSKRPR-----VPKRAQALE 2498
2315 -QPLSRPQSPGRNISPGRNGIS--PPNKLSQLPRTSSPSTASTKSGSGKM-SYTSRG 2370
2499 LAGLKGPIQLGHTDESASDKGPTQPCNSLOPEQVDSQSSPRPRPTRRGKVEADEEPS 2558
2371 RQMSQQNLTKOTGLSKNASSIPRESASKGLNMNGNGANK-KVLSRMSSTKSGSES 2429
2559 AVKTVSTSRQTRSRKVPKPEIG-----NNGTQVSKASIKQTLDTVAKTGVS-- 2604
2430 DRSERPVLVQSTFIKEAPSTLRRKLEESASFESLSPSRSPASPTESQAQTPVLSPLP 2489
2605 -----RQURTHKGWGSTLLKLGDSKETQISDHSE----- 2636
2490 DMSLSTHSSVQAGWRKLPNLPNLTIEYNDGRPAKHDIARSHSPSLPILNRSQTKR 2549
2637 --KLAHDSILKSTQOQKQPSVKPL-----RCCR--VLRASKEVPEKVLVD-----TR 2681
2550 EHSK-HSSILPRVSTWRTGSSSILSASSESSEKAKSE--DE----- 2589
2682 DHATQSKSNPLSPRKRKADGSIWTRALRSLAPKQEADEKPVPEKRAASSKRYVS 2741
2590 -----KHNASISGTSQKSNQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAESKTL 2643
2742 PEPVMCHLKI VSNKLESVEQVS---TVMKTEMEAKREN-----PVTQDQNSRYRKTN 2794
2644 IYQMAPAV-SKTEDVWVRIEDCFINPRSGRSTGNTPPVIDSVSEKAPNIKQSKNOA 2702
2795 VKQPRKFDASAENVGK-----KNEKTMKTASQET-----ELQNP-----DDGA 2834
2703 KQNVNGSVPMRTVGLNRLNSFIQVADPQKQTEIKPGQNNPVPVSETNESSIVERTPF 2762

Db 2835 KKTSRQGVGKRTCLSR-----GTTEMP---QCEABE----- 2866
QY 2763 SSSSSSHSGSPGVAARVTFPNYNPSPRKSADSTARSQI-----PTFVANNTKRD 2817
Db 2867 -----KTSKPAEIL--IKPOBEKGVSGSDVRCLSRKRTRVALDSEPKPRVTRGTXKD 2918
QY 2818 SKT 2820
Db 2919 AKT 2921
RESULT 15
T20532
Hypothetical protein F07A11.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20532; T27777
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20532
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2722 <WIL>
A:Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CBSP:F07A11.6b
A:Experimental source: clone F07A11
R:Gajadecy, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27777
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2722 <WIL>
A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CBSP:F07A11.6b
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CBSP:F07A11.6b
A:Map position: 2
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 1286/3; 15
Query Match 2.9%; Score 417.5; DB 2; Length 2722;
Best Local Similarity 18.6%; Pred. No. 3.8e-09;
Matches 460; Conservative 324; Mismatches 899; Indels 789; Gaps 107;
QY 811 SDNFTGNMTVLSYLVNTTVLPSSSSSRGSLDSSRSSEKDRSLERERIGLGNHYHPATENP 870
Db 357 SDRENL-----YFINRKENIEVAARSSSPSKSENQ-----GSSSP----- 394
QY 871 GTSKRGGLQISTTAAQIAKWEVSAIHTSQEDR---SSGTTTELHCVTERNALRRSSA 927
Db 395 --SSSRDRQNLHDLQTRSSVEH---HTNQDOENNASGDS-----SSDSSEEGSSS 443
QY 928 AHTHSNTVNTFKSNSNRCTCMPYAKLEYKRSSNDLSNVSSSDCYGKRGQMKPSEISYS 987
Db 444 SNEQSDQNDVDEDDDDVVS-----EKKR--HEPEGKSSSPGNGHRDESNG--DKDH 493
QY 988 EDDSKFCYSQYQPADIAHKIHSANHMDDNDGEL-----DTPINYSKYDQNLNQRQS 1042
Db 494 EDSSEFP-----SQPSTSSH--HETSHSPEKDSEAYQSRFSPLNY-----QS 534
QY 1043 PSQNERWARPKHILEDIKQS-EQOQRNOSTTTPVYTESTDD-----KHLKFOHPFGQ 1095
Db 535 QSPGYEFLESK-----ELKQEFSTTSASSDLEDMENPDPDLTRMLERHWRPID- 588
QY 1096 QECVSPYRSGANGSETNRVGSNHNQVNSQSL-----CQEDDYED-----DK 1139
Db 589 ---VSSF-----VNRIDEIVELNQAARASVEKFTGRFPFKCANNDEVLSIQKIVFHE 636
QY 1140 PTNYSERYSEBEQHEEERPTNYSIKYNEEKHVDQIDYSLKYATDIPSSQKQSPFSK 1199
Db 637 PRDY--YYENPCSELEVRIDW-----RKLSDTADLD-----DFRAIDSKELGRDQ 681

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:21 ; Search time 32.5 Seconds
(without alignments)
4554.937 Million cell updates/sec

Title: US-09-442-489F-7

Perfect score: 14566

Sequence: 1 MARASYDQLLKQVEALKMEN.....SSGTSQPKRHSGSYLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14566	100.0	2843	1 APC_HUMAN	P25054 homo sapien
2	13198.5	90.6	2842	1 APC_RAT	P70478 rattus norv
3	13178	90.5	2845	1 APC_MOUSE	Q61315 mus musculu
4	476	3.3	2774	1 MAPA_RAT	P34926 rattus norv
5	474	3.3	5560	1 SPEN_DROME	Q8sx83 drosophila
6	452	3.1	3664	1 MINT_HUMAN	Q96t58 homo sapien
7	428.5	2.9	3644	1 MINT_MOUSE	Q42504 mus musculu
8	404	2.8	2688	1 ZBP1_MOUSE	Q31172 mus musculu
9	401.5	2.8	2805	1 MAPA_HUMAN	P78559 homo sapien
10	399	2.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
11	391.5	2.7	3562	1 PCV1_CHICK	Q90953 gallus gall
12	390	2.7	5085	1 PCLO_RAT	Q9jks6 rattus norv
13	389.5	2.7	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
14	380.5	2.6	4911	1 MLF3_HUMAN	Q8nez4 homo sapien
15	379.5	2.6	3210	1 CENF_HUMAN	P49454 homo sapien
16	375.5	2.6	2459	1 MAPB_RAT	P15205 rattus norv
17	374.5	2.6	4377	1 ANK3_HUMAN	Q12955 homo sapien
18	374	2.6	3256	1 KI67_HUMAN	P46013 homo sapien
19	372	2.6	5120	1 PCLO_CHICK	Q9pu36 gallus gall
20	371.5	2.6	1581	1 PRB_HUMAN	Q95648 h peroxisom
21	370.5	2.5	3969	1 HRX_HUMAN	Q31164 homo sapien
22	370	2.5	2453	1 NCR1_MOUSE	Q60974 mus musculu
23	365	2.5	1411	1 TCOF_HUMAN	Q33428 homo sapien
24	365	2.5	2440	1 NCR1_HUMAN	Q75376 homo sapien
25	363	2.5	2468	1 MAPB_HUMAN	P46821 homo sapien
26	363	2.5	3259	1 G0B1_HUMAN	Q14789 homo sapien
27	362.5	2.5	2663	1 CENE_HUMAN	Q02224 homo sapien
28	362	2.5	2492	1 ATRX_PANTR	Q7yqm4 pan troglod
29	361	2.5	7389	1 PFAL_MOUSE	Q9aiz6 mus musculu
30	359	2.5	5038	1 PCLO_MOUSE	Q9gyx7 mus musculu
31	358.5	2.5	6632	1 UN89_CAEEL	Q01761 caenorhabdi
32	357	2.5	2464	1 MAPB_MOUSE	P14873 mus musculu
33	356.5	2.4	1972	1 P531_HUMAN	Q12888 homo sapien

ALIGNMENTS

```

RESULT 1
APC_HUMAN
ID APC_HUMAN STANDARD; PRT: 2843 AA.
AC P25054; Q15162; Q15163;
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC OR DP2.5 (Human)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335210; PubMed=1651562;
RA Kinzler L., Smith K.J., Vogelstein B., Bryan T.M.,
RA Levy D.B., Hedger P., McKechnie D.,
RA Fennell R., Markham A., Groffen J., Boguski M.S., Altschul S.F.,
RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,
RA "Identification of FAP locus genes from chromosome 5q21."
RL Science 253:661-665(1991).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=91330307; PubMed=1678319;
RA Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,
RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
RA Sargeant L., Kravcho K., Wolff E., Burt R., Hughes J.P.,
RA Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,
RA Abderrahim H., Cohen D., Leppert M., White R.;
RA "Identification of deletion mutations and three new genes at the
RA familial polyposis locus."
RL Cell 66:601-613(1991).
RN [3]
RP ASSOCIATION WITH CATENINS.
RX MEDLINE=94082295; PubMed=8259519;
RA Su L.-K., Vogelstein B., Kinzler K.W.;
RA "Association of the APC tumor suppressor protein with catenins."
RL Science 262:1734-1737(1993).
RN [4]
RP DISEASE.
RX MEDLINE=95174843; PubMed=7661930;
RA Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,
RA Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,
RA Wood P.A., Taqi F., Booker S.V., Petersen G.M., Offerhaus G.J.A.,
RA Tersmette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.;
RA "The molecular basis of Turcot's syndrome."
RL New Engl. J. Med. 332:839-847(1995).
RN [5]
RP DISEASE.
RX MEDLINE=97094176; PubMed=8940264;
RA Eccles D.M., van der Luijt R.B., Breukel C., Bullman H., Bunyan D.,
RA Fisher A., Barber J., du Boulay C., Primrose J., Fodde R.;
RA "Hereditary desmoid disease due to a frameshift mutation at codon 1924
RA of the APC gene."
RT

```

P08640 saccharomyc
P55200 mus musculu
P30414 homo sapien
Q03661 saccharomyc
Q9y618 h nuclear r
P30415 mus musculu
P46100 homo sapien
Q9wu42 mus musculu
Q95923 rattus norv
Q9nzw4 homo sapien
Q7yqm3 pongo pygma
P47035 saccharomyc

- Am. J. Hum. Genet. 59:1193-1201(1996).
[6]
RN DISEASE.
RP MEDLINE=20243021; PubMed=10782927;
RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
RA Rodde R., Almon B., Bapat B.;
RT "A germline mutation at the extreme 3-prime end of the APC gene
RT results in a severe desmoid phenotype and is associated with
RT overexpression of beta-catenin in the desmoid tumor.";
RL Clin. Genet. 57:205-212(2000).
[7]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 2-55.
RP MEDLINE=20384842; PubMed=10926498;
RA Day C.L., Alber T.;
RT "Crystal structure of the amino-terminal coiled-coil domain of the
RT APC tumor suppressor.";
RL J. Mol. Biol. 301:147-156(2000).
[8]
RN X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 1021-1035 IN COMPLEX WITH
RP CTNNB1.
RX MEDLINE=21564054; PubMed=11707392;
RA Ekiof Spink K., Fridman S.G., Weiss W.I.;
RT "Molecular mechanisms of beta-catenin recognition by adenomatous
RT polyposis coli revealed by the structure of an APC-beta-catenin
RT complex.";
RL EMBO J. 20:6203-6212(2001).
[9]
RN X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
RP AXIN.
RX MEDLINE=20271867; PubMed=10811618;
RA Spink K.E., Polakis P., Weiss W.I.;
RT "Structural basis of the axin-adenomatous polyposis coli
RT interaction.";
RL EMBO J. 19:2270-2279(2000).
[10]
RN REVIEW ON VARIANTS.
RP MEDLINE=94154728; PubMed=8111410;
RA Nagase H., Nakamura Y.;
RT "Mutations of the APC (adenomatous polyposis coli) gene.";
RL Hum. Mutat. 2:425-434(1993).
[11]
RN VARIANTS FAP.
RX MEDLINE=91335211; PubMed=1651563;
RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
RT patients.";
RL Science 253:665-669(1991).
[12]
RN VARIANTS FAP.
RX MEDLINE=93265030; PubMed=1338904;
RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
RA Miki Y., Mori T., Nakamura Y.;
RT "Somatic mutations of the APC gene in colorectal tumors: mutation
RT cluster region in the APC gene.";
RL Hum. Mol. Genet. 1:229-233(1992).
[13]
RN VARIANTS FAP.
RX MEDLINE=93244793; PubMed=1338691;
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
RA Nakamura Y., Horii A.;
RT "Somatic mutation of the APC gene in gastric cancer: frequent
RT mutations in very well differentiated adenocarcinoma and signet-ring
RT cell carcinoma.";
RL Hum. Mol. Genet. 1:559-563(1992).
[14]
RN VARIANT FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND
RP SER-2502.
RX MEDLINE=93250848; PubMed=1338764;
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
Baba S., Nakamura Y.;
RT "Screening for germ-line mutations in familial adenomatous polyposis
RT patients: 61 new patients and a summary of 150 unrelated patients.";
RL Hum. Mutat. 1:467-473(1992).
[15]
RN VARIANT FAP TRP-99.
RX TISSUE=Peripheral blood lymphocytes;
RP MEDLINE=95134544; PubMed=7831149;
RA Dobble Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,
RA Mueller A., Mueller H., Scott R.J.;
RT "Mutational analysis of the first 14 exons of the adenomatous
RT polyposis coli (APC) gene.";
RL Eur. J. Cancer 30A:1709-1713(1994).
[16]
RN VARIANT FAP GLY-722.
RP MEDLINE=95135430; PubMed=7833931;
RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Mareni C., Quanti G.;
RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
RT FAP patients.";
RL Hum. Mol. Genet. 3:1687-1688(1994).
[17]
RN ERRATUM.
RP Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Mareni C., Quanti G.;
RL Hum. Mol. Genet. 3:1918-1918(1994).
[18]
RN VARIANT FAP ILE-171.
RX MEDLINE=97144176; PubMed=8990002;
RA van der Luijt R.B., Meera Khan E., Vassen H.F.A., Tops C.M.J.,
RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
RA Plug R.J., Griffioen G., Fodde R.;
RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
RT familial adenomatous polyposis: 67 germline mutations identified by
RT DGGE, PTT, and southern analysis.";
RL Hum. Mutat. 9:7-16(1997).
[19]
RN VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.
RX MEDLINE=98080146; PubMed=9419979;
RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
RA Fukutome A., Tomiyama J., Chuganji Y., Monoi M., Utsunomiya J.;
RT "Drastic genetic instability of tumors and normal tissues in Turcot
RT syndrome.";
RL Oncogene 15:2877-2881(1997).
[20]
RN VARIANT LYS-1307.
RX MEDLINE=98400248; PubMed=9731522;
RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
RT "The APC I1307K allele and breast cancer risk.";
RL Nat. Genet. 20:13-14(1998).
[21]
RN VARIANTS LYS-1307 AND GLN-1317.
RP TISSUE=Peripheral blood;
RX MEDLINE=98393712; PubMed=9724771;
RA Frayling I.M., Beck N.B., Ilyas M., Dove-Edwin I., Goodman P.,
RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
RT "The APC variants I1307K and E1317Q are associated with colorectal
RT tumors, but not always with a family history.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
[22]
RN VARIANT LYS-1307.
RX MEDLINE=98400259; PubMed=9731533;
RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
RT "The APC I1307K allele and cancer risk in a community-based study of
RT Ashkenazi Jews.";
RL Nat. Genet. 20:62-65(1998).

Query Match 100.0%; Score 14566; DB 1; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAYDQLLKQVEALKMENSRLROLEDNSNHLTKLETEASNMEKVLKQLQGSIEDAM 60
 Db 1 MAAAYDQLLKQVEALKMENSRLROLEDNSNHLTKLETEASNMEKVLKQLQGSIEDAM 60

QY 61 ASSGOIDLLERLKELEKLNLOSSNPPGKLRKMSLRSGREGSVSRGECFVPVPMGSPPR 120
 Db 61 ASSGOIDLLERLKELEKLNLOSSNPPGKLRKMSLRSGREGSVSRGECFVPVPMGSPPR 120

QY 121 RGTVNGSRESTGYLBELEKERSLLMLADLDKBEKEDWYAOQLNLTAKIDSLPTENPSL 180
 Db 121 RGTVNGSRESTGYLBELEKERSLLMLADLDKBEKEDWYAOQLNLTAKIDSLPTENPSL 180

QY 181 QTDWTRQLEYEARQIRVAMEEQLGTCQDMKRAQRRIARQQTEKOTLRIRQLQSOAT 240
 Db 181 QTDWTRQLEYEARQIRVAMEEQLGTCQDMKRAQRRIARQQTEKOTLRIRQLQSOAT 240

QY 241 EAEERSQNKHETGSHDAERQEGQVGEINMATSNGGQSTTRMDHETASVLSSTHSA 300
 Db 241 EAEERSQNKHETGSHDAERQEGQVGEINMATSNGGQSTTRMDHETASVLSSTHSA 300

QY 301 PRRLTSHLGTKEVWYVLSMLGTHDXDDMERTLLAMSSQDSCISMRSQSGCLPLLQLL 360
 Db 301 PRRLTSHLGTKEVWYVLSMLGTHDXDDMERTLLAMSSQDSCISMRSQSGCLPLLQLL 360

QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHSQDDPKRGREIRVLHLLQIRAYCETC 420
 Db 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHSQDDPKRGREIRVLHLLQIRAYCETC 420

QY 421 WEQBAEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEBHRHAMNELGLQIAELLQ 480
 Db 421 WEQBAEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEBHRHAMNELGLQIAELLQ 480

QY 481 VDCMYGHTNDHYSITLRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSEEDL 540
 Db 481 VDCMYGHTNDHYSITLRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSEEDL 540

QY 541 QQVIASVLRNLNLSWRADVNSKKTLEVGSVKALMECALVEKKESTLKSVLKALWNLASHCT 600
 Db 541 QQVIASVLRNLNLSWRADVNSKKTLEVGSVKALMECALVEKKESTLKSVLKALWNLASHCT 600

QY 601 ENKADICAVDQALAPLVGTLTYRSQNTNLAIIESGGILNRVNSLLIATNEDHRQILRENN 660
 Db 601 ENKADICAVDQALAPLVGTLTYRSQNTNLAIIESGGILNRVNSLLIATNEDHRQILRENN 660

QY 661 CLQTLLOHLKSHSLTIIVSNACGTLWNLNLSARPKQDEALWDMGAVSMLKNLIHSHKXMIAM 720
 Db 661 CLQTLLOHLKSHSLTIIVSNACGTLWNLNLSARPKQDEALWDMGAVSMLKNLIHSHKXMIAM 720

QY 721 GSAALRLNLMANRPKAYKXANIMSPGSSLPVSLHVRKQKALEALDAOHLSETFDNIDNLS 780
 Db 721 GSAALRLNLMANRPKAYKXANIMSPGSSLPVSLHVRKQKALEALDAOHLSETFDNIDNLS 780

QY 781 PKASHRSQKORHQSILYGVDFPTNRHDDNRDNFTNGMTVLSPLYLNTVLPSSSSSRGS 840
 Db 781 PKASHRSQKORHQSILYGVDFPTNRHDDNRDNFTNGMTVLSPLYLNTVLPSSSSSRGS 840

QY 841 LDSSRSEKDRSLERERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSAIHTS 900
 Db 841 LDSSRSEKDRSLERERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSAIHTS 900

QY 901 QEDRSSGTTLHCVTDERNALRSSAAHTSNYNTFKSENSNRTCSMPYAKLEYKRSS 960
 Db 901 QEDRSSGTTLHCVTDERNALRSSAAHTSNYNTFKSENSNRTCSMPYAKLEYKRSS 960

QY 961 NDSLNSVSSSDGYKRGQMKPSIESYEDDSEKSCSYGOYPADLAHKITHSANHMDNDGE 1020
 Db 961 NDSLNSVSSSDGYKRGQMKPSIESYEDDSEKSCSYGOYPADLAHKITHSANHMDNDGE 1020

QY 1021 LDTPIYSLKYSDEQLNSGRQSPSONERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
 Db 1021 LDTPIYSLKYSDEQLNSGRQSPSONERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080

QY 1081 STDDXHLKFPQHPHQECVSPYASRGANGSETNRVSGNHGINQVNSOSLQCOEDDYEDDKP 1140
 Db 1081 STDDXHLKFPQHPHQECVSPYASRGANGSETNRVSGNHGINQVNSOSLQCOEDDYEDDKP 1140

QY 1141 TNSYERYSEEEQHEEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSPKS 1200
 Db 1141 TNSYERYSEEEQHEEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSPKS 1200

QY 1201 SSGQSSKTEHWSSESSENTSTPSSNAKRONOLHPSSAQSRGQPKAATCKVSSINOBTIQ 1260
 Db 1201 SSGQSSKTEHWSSESSENTSTPSSNAKRONOLHPSSAQSRGQPKAATCKVSSINOBTIQ 1260

QY 1261 TYCVEDTPICFSRCSSLSLSSAEDBIGNCQTTQZADSAANTLQIAIEKIKIGTRSAEDPV 1320
 Db 1261 TYCVEDTPICFSRCSSLSLSSAEDBIGNCQTTQZADSAANTLQIAIEKIKIGTRSAEDPV 1320

QY 1321 SEVPAYSOHPTKSSRLQSSLSSESAHKAHVPFSSGAKSPKSGAQTPKSPPEHYVQET 1380
 Db 1321 SEVPAYSOHPTKSSRLQSSLSSESAHKAHVPFSSGAKSPKSGAQTPKSPPEHYVQET 1380

QY 1381 PLMFSRCTSVSSLDSPESRSIASVQSEPCSGMVSGIISPSDLDPSPGQTMPPSRSKTTP 1440
 Db 1381 PLMFSRCTSVSSLDSPESRSIASVQSEPCSGMVSGIISPSDLDPSPGQTMPPSRSKTTP 1440

QY 1441 PPQTQATKREVPKNAKPAETAEKESGPKQAAVNAAVQVRQVLPDADTLHFAESTPDGF 1500
 Db 1441 PPQTQATKREVPKNAKPAETAEKESGPKQAAVNAAVQVRQVLPDADTLHFAESTPDGF 1500

QY 1501 SCSSLSALSALDLPFTQKVELRIMPVQENDNGNTESEQPKESNENKEKEAKTIIDSE 1560
 Db 1501 SCSSLSALSALDLPFTQKVELRIMPVQENDNGNTESEQPKESNENKEKEAKTIIDSE 1560

QY 1561 KOLLDDSDDDDIIEBEECIIISAMPTKSSRAKKAQATASKLPPPVARKPSQLPVYKLLPS 1620
 Db 1561 KOLLDDSDDDDIIEBEECIIISAMPTKSSRAKKAQATASKLPPPVARKPSQLPVYKLLPS 1620

QY 1621 QNRLQPKQKVSFTPGDMPRVYCVETGPIINFSTATSLDLTIESPNELAAEGVGGQA 1680
 Db 1621 QNRLQPKQKVSFTPGDMPRVYCVETGPIINFSTATSLDLTIESPNELAAEGVGGQA 1680

QY 1681 SGFEKXRTIPTEGRSTDEAQQGKTSVTIPELDDNKAEEGDILAEICINSAMPKGSHKP 1740
 Db 1681 SGFEKXRTIPTEGRSTDEAQQGKTSVTIPELDDNKAEEGDILAEICINSAMPKGSHKP 1740

QY 1741 FRVKKIMDQVQASASSAPNKNQLDGKKKPTSPVKPIPONTEYTRVRKNADSKNNLN 1800
 Db 1741 FRVKKIMDQVQASASSAPNKNQLDGKKKPTSPVKPIPONTEYTRVRKNADSKNNLN 1800

QY 1801 AERVFSDNKSQKQNLKNNKDFNDKLPNNEDVRGSAFDPGPHHTPIEGTPICFSRND 1860
 Db 1801 AERVFSDNKSQKQNLKNNKDFNDKLPNNEDVRGSAFDPGPHHTPIEGTPICFSRND 1860

QY 1861 SLSSLDPDDDDVDLSREKAEKAKENKSEAKVTSHTELTISNQQSANKTQAIKQPINR 1920
 Db 1861 SLSSLDPDDDDVDLSREKAEKAKENKSEAKVTSHTELTISNQQSANKTQAIKQPINR 1920

QY 1921 GQPKPILOKQSTFPQSSKDIIDRGAAATDEKLQNAFENTPVCFSNSLSLSDIDQENN 1980
 Db 1921 GQPKPILOKQSTFPQSSKDIIDRGAAATDEKLQNAFENTPVCFSNSLSLSDIDQENN 1980

QY 1981 NKNEPIKETEPDPSQGEPSKPAQAGYAPKSFHVEDTPVCFSRNSLSLSIDSEDDLQ 2040
 Db 1981 NKNEPIKETEPDPSQGEPSKPAQAGYAPKSFHVEDTPVCFSRNSLSLSIDSEDDLQ 2040

QY 2041 ECISAMPKKKPSRLKGNKEKHSRNMKGILGEDLTLDKDIQRPDSEHGLSPSENFD 2100
 Db 2041 ECISAMPKKKPSRLKGNKEKHSRNMKGILGEDLTLDKDIQRPDSEHGLSPSENFD 2100

QY 2101 WKAIORGANSIVSSLHQAAAAACLRQASSDSLSLSKSGISLGSFPFLHTPDQEKPT 2160

Db 2101 WKAIQEGANSIVSSLSHQAACAAACLSRQSSSDSLSLKSIGSLGSPHLLTPDQEKPT 2160
 QY 2161 SNKGPRLPGEKSTLETTKIESEKIGKGVKYSLLTGKVRNSSETSGOMKOPLOAN 2220
 Db 2161 SNKGPRLPGEKSTLETTKIESEKIGKGVKYSLLTGKVRNSSETSGOMKOPLOAN 2220
 QY 2221 MPSISGRMTWHPGVNRSSTSPVSKGPKLPKTPAKSPSEGGTATTPRGAKSPVKS 2280
 Db 2221 MPSISGRMTWHPGVNRSSTSPVSKGPKLPKTPAKSPSEGGTATTPRGAKSPVKS 2280
 QY 2281 ELSFVARQTSQIGSSKAPRSRSDSTSPRAQPLSRPIOSPGRNLSIPRNGISPN 2340
 Db 2281 ELSFVARQTSQIGSSKAPRSRSDSTSPRAQPLSRPIOSPGRNLSIPRNGISPN 2340
 QY 2341 KLSQLPRTSPSTASTKSSGSGKMSYTPSPQMSQQNLTKQGLSKNASSIPRSESAGK 2400
 Db 2341 KLSQLPRTSPSTASTKSSGSGKMSYTPSPQMSQQNLTKQGLSKNASSIPRSESAGK 2400
 QY 2401 LQNMNNGANKVELSRMSTKSSGSESDRSPRPVLRQSTFIKEAPSPTLRKLBSA 2460
 Db 2401 LQNMNNGANKVELSRMSTKSSGSESDRSPRPVLRQSTFIKEAPSPTLRKLBSA 2460
 QY 2461 SFESLSPSPSPASPTQSOATPVLSPLDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Db 2461 SFESLSPSPSPASPTQSOATPVLSPLDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 QY 2521 RPAKRDHDIARSHESPSRLPINRSGTWKREHSHSSLPVSTWRTTSGSSSILSASSBS 2580
 Db 2521 RPAKRDHDIARSHESPSRLPINRSGTWKREHSHSSLPVSTWRTTSGSSSILSASSBS 2580
 QY 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWKIKENEFSPNNSQTVSSGATNGAES 2640
 Db 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWKIKENEFSPNNSQTVSSGATNGAES 2640
 QY 2641 KTLIIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKN 2700
 QY 2701 QAKQNGVGVMPRTVGLNRLNSFTQVDAPDQKTEIKPQNNPVVSETNESSIVERT 2760
 Db 2701 QAKQNGVGVMPRTVGLNRLNSFTQVDAPDQKTEIKPQNNPVVSETNESSIVERT 2760
 QY 2761 PFSSSSSKHSSPSGTVAARVTFPNYNSPRKSADSTSGARPSQIPTPVNNNTKRDST 2820
 Db 2761 PFSSSSSKHSSPSGTVAARVTFPNYNSPRKSADSTSGARPSQIPTPVNNNTKRDST 2820
 QY 2821 DSTESSGTQSPKRGSGSYLTVSV 2843
 Db 2821 DSTESSGTQSPKRGSGSYLTVSV 2843

RESULT 2
 APC_RAT
 ID APC_RAT STANDARD; PRT; 2842 AA.
 AC P70478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID:10116;
 EN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
 RA Sugimura T., Nagao M.;
 RT "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
 RL Mamm. Genome 6:746-748(1995).

RN [2]
 RP MUTAGENESIS.
 RC STRAIN=Sprague-Dawley, and Fischer 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
 RA Weisburger J.H., Sugimura T., Nagao M.;
 RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
 tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
 CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 and participates in Wnt signaling. APC activity is correlated with
 its phosphorylation state (By similarity).
 CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
 axin (By similarity).
 CC -!- PTM: Phosphorylated by GSK3B (By similarity).
 CC -!- SIMILARITY: Contains 7 ARM repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D38629; BAA07609.1; ..
 DR HSP; Q02248; 3BCT.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg; 4.
 DR SMART; SMO0185; ARM; 5.
 DR PROSITE; PS0176; ARM_REPEAT; 1.
 KW Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
 Repeat.
 FT DOMAIN 1 728 LEU-RICH.
 FT DOMAIN 125 260 COILED COIL (POTENTIAL).
 FT REPEAT 451 493 COILED COIL (POTENTIAL).
 FT REPEAT 503 545 ARM 1.
 FT REPEAT 546 589 ARM 2.
 FT REPEAT 590 636 ARM 3.
 FT REPEAT 637 681 ARM 4.
 FT REPEAT 682 723 ARM 5.
 FT REPEAT 724 765 ARM 6.
 FT REPEAT 739 783 ARM 7.
 FT DOMAIN 1130 1155 SER-RICH.
 FT DOMAIN 1356 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 ASP/GLU-RICH (ACIDIC).
 FT MUTAGEN 523 523 HIGHLY CHARGED.
 FT SEQUENCE 2842 AA; 310530 MW; 3CB2E8A8A3E8F47 CRC64;
 SQ
 Query Match 90.6%; Score 13198.5; DB 1; Length 2842;
 Best Local Similarity 90.0%; Pred. No. 0;
 Matches 2563; Conservative 121; Mismatches 154; Indels 9; Gaps 8;
 QY 1 MAASVDQLKQVEALKMENSRLQELFEDNSNHLTKLETSANNKEVYLKQLQGSIEDEAM 60
 Db 1 MAASVDQLKQVEALKMENSRLQELFEDNSNHLTKLETSANNKEVYLKQLQGSIEDEAM 60
 QY 61 ASSGQIDLLERIKELNLDSSNFPFGVKLRKMSLSYSGRSVSRSGECSPVPMGSPFR 120
 Db 61 -TSQIDLLERLKEFNLD-SNFPFGVKLRKMSLSYSGRSVSRSGECSPVPMGSPFR 118
 QY 121 RGFVNSRESYGLLEKEKSLLLADLKEEKDQVYAOQLNLTREIDSLPTENFSL 180
 Db 119 RAFVNSRESYGLLEKEKSLLLADLKEEKDQVYAOQLNLTREIDSLPTENFSL 178
 QY 181 QDTMTFRQLEYEARQIRVAMEEQLGTCDMEKRAQRRIARIQQIEKDIRLROLQSOAT 240
 Db 179 QDTMTFRQLEYEARQIRVAMEEQLGTCDMEKRAQRRIARIQQIEKDIRLROLQSOAA 238
 QY 241 EAERSSQNHGTGSHDAERQNEGVGVNMTSGNGGSGTTRMDHETASVLSSTHSA 300

Db 239 EAERSQKHETASHEABERQLEGGQVAESNLATSGGSSAAARVDHETAGVLSSSGTHSA 298
Qy 301 PRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSQSDSCISMRSQSGCLPILLIQL 360
Db 299 PRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSQSDSCISMRSQSGCLPILLIQL 358
Qy 361 HGNDDKSVLLGNSRSGKEARASAAALHNIHISQDDDKRGRREIRVLLHLLBOIRAYCETC 420
Db 359 HGNDDKSVLLGNSRSGKEARASAAALHNIHISQDDDKRGRREIRVLLHLLBOIRAYCETC 418
Qy 421 WEMQBAHEPGMDQKNPMPAPVEHQCIPAVCVLMKLSFDEEHRHAMNELGGLOATAELLO 480
Db 419 WEMQBAHEPGMDQKNPMPAPVEHQCIPAVCVLMKLSFDEEHRHAMNELGGLOATAELLO 478
Qy 481 VDCMYGLTNDHYSITLRYAGMALTNLTFGQVANKATLCMKGCMRALVAQLKSESDL 540
Db 479 VDCMYGLTNDHYSITLRYAGMALTNLTFGQVANKATLCMKGCMRALVAQLKSESDL 538
Qy 541 QOVIASVLNLSWRADVNSKTLREVGSKALMECALEVKESTLKSVLNLSAHCT 600
Db 539 QOVIASVLNLSWRADVNSKTLREVGSKALMECALEVKESTLKSVLNLSAHCT 598
Qy 601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILRNVSLLIATNEDHRQLRENN 660
Db 599 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILRNVSLLIATNEDHRQLRENN 658
Qy 661 CLQTLQHLKSHSLTIVSNACOTLNLNLSARNPKDQEAALWDGAVNMLKVLHSHKXMIAM 720
Db 659 CLQTLQHLKSHSLTIVSNACOTLNLNLSARNPKDQEAALWDGAVNMLKVLHSHKXMIAM 718
Qy 721 GSAAALNLMANRPKAKYKDIANIMSGSLPSLHVKKQKALAEALDAOHLSETFNDLNS 780
Db 719 GSAAALNLMANRPKAKYKDIANIMSGSLPSLHVKKQKALAEALDAOHLSETFNDLNS 778
Qy 781 PKASHRSQRKHQSUYGVYFDTNRHDDNRDNFNNTGNMVLSPYLNLTVPSSSSSRGS 840
Db 779 PKASHRSQRKHQSUYGVYFDTNRHDDNRDNFNNTGNMVLSPYLNLTVPSSSSSRGS 838
Qy 841 LDSSRSKDRSLERBERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSALHTS 900
Db 839 LDSSRSKDRSLERBERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSALHTS 898
Qy 901 QEDRSSGTTTELCHVTDERNALRRSSAAHTSNYNTFTKSENSNRCTCMYPKALEYKRSS 960
Db 899 QEDRSSGTTTELCHVTDERNALRRSSAAHTSNYNTFTKSENSNRCTCMYPKALEYKRSS 958
Qy 961 NDLSNVSSDGYGKRGOMKPSIERYSDDSKFCYQYPADLAHKHISANHMDNDGE 1020
Db 959 NDLSNVSSDGYGKRGOMKPSIESYSDDBGKFCYQYPADLAHKHISANHMDNDGE 1018
Qy 1021 LDTPIYSLKYSDEQLNSGRQSPSONERWARPKHIEDEIKQSEORQORNOSTTYPVYTE 1080
Db 1019 LDTPIYSLKYSDEQLNSGRQSPSONERWARPKHIEDEIKQSEORQORNOSTTYPVYTE 1078
Qy 1081 STDDKHLKFPQHFQOQECVPSYRSGANGSETNRVGSNHGNQNVOSLCOEDDYEDDKP 1140
Db 1079 NTDDKHLKFPQHFQOQECVPSYRSGANGSETNRVGSNHGNQNVOSLCOEDDYEDDKP 1138
Qy 1141 TNYSERYSEREEHREERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSFSEKS 1200
Db 1139 TNYSERYSEREEHREERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSFSEKS 1198
Qy 1201 SSGQSKTEHSSSSSENTSTPSSNAKRONQLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Db 1199 PSVQTKTEHNSPSSSEAAASAPSSNAKRSQSLHPSSAQ-RNQTPKGTACKVPSINQETMQ 1257
Qy 1261 TYCVEDTPICSRCSLSLSAEDIEGNCOTTOEADSANLQIAEIKELIGTSAEDPV 1320
Db 1258 TYCVEDTPICSRCSLSLSAEDIEGNCOTTOEADSANLQIAEIKELIGTSAEDPV 1317
Qy 1321 SEVPAVSQHPRTKSRLQSGSLSSSARHKAHVEFPSSGAKSPSKGAOTPKPPEHYVQET 1380
Db 1318 SDVPAVSQSTRKPSRLQASGLASESARHKAHVEFPSSGAKSPSKGAOTPKPPEHYVQET 1377

Qy 1381 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGOTMPSRSKT-- 1438
Db 1378 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGOTMPSRSKTTP 1437
Qy 1439 PPPPQTAQTKREVVKNAKPTAKRBSGPKQAAVANAQVQVLPDADTLLHFATESTPD 1498
Db 1438 PPPPQVQTKREVVKNAKPTAKRBSGPKQTAVAQAQVQVLPDADTLLHFATESTPD 1497
Qy 1499 GFSCSSLSALSLEDEPFIQKDVLRMPVQENDNGNETESEOPKSENQKEAEKTIID 1558
Db 1498 GFSCSSLSALSLEDEPFIQKDVLRMPVQENDNGNETESEOPKSENQKEAEKTIID 1556
Qy 1559 SEKDLDDSDDDIEILEECIIISAMPTKSRKAKKPAQTASKLPPPVARXPSQLPVYKLL 1618
Db 1557 SEKDLDDSDDDIEILEECIIISAMPTKSRKAKKPAQTASKLPPPVARXPSQLPVYKLL 1616
Qy 1619 PSQNLQOPKHVFTPGDDMPRVYCVBGTPIINSTATSLSDLIIESPPNELAAGEVRGG 1678
Db 1617 PSQNLQOPKHVFTPGDDMPRVYCVBGTPIINSTATSLSDLIIESPPNELAAGEVRGG 1676
Qy 1679 AQGEFEKRTIITEGRSTDEAOGGKTSSVTIPELDDNKAEEDGIIAECINSAMPKXSH 1738
Db 1677 AQGEFEKRTIITEGRSTDEAOGGKTSSVTIPELDDNKAEEDGIIAECINSAMPKXSH 1736
Qy 1739 KPRVKKIMQVQOASASSAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKN 1798
Db 1737 KPRVKKIMQVQOASASSAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKN 1796
Qy 1799 LNAERFSDNKNKQNLKNNKDFDKLPNNEDRVGRSFAFDSPHHYTPIEGTVCFSR 1858
Db 1797 VNTTEFSDNKNKQNLKNNKDFDKLPNNEDRVGRSFAFDSPHHYTPIEGTVCFSR 1856
Qy 1859 NDLSLSDPDDDDVLSREKAEELKAKENKESAKVTSHTELTSNQOSANKTOAIKQPI 1918
Db 1857 NDLSLSDPDDDDVLSREKAEELKAKENKESAKVTSHTELTSNQOSANKTOAIKQPI 1916
Qy 1919 NRQOPKILQKQSTFPQSSKDIIDRGAATDEKLONFAIENTPVCFSHNSLSLSLSDOE 1978
Db 1917 NRQOPKILQKQSTFPQSSKDIIDRGAATDEKLONFAIENTPVCFSHNSLSLSLSDOE 1976
Qy 1979 -NNKNNEPIKETEPPDPSQGEPSKPOASGYAPASFHVEDTPVCFSRNSLSLSLSDOE 2037
Db 1977 -NNKNNEPIKETEPPDPSQGEPSKPOASGYAPASFHVEDTPVCFSRNSLSLSLSDOE 2036
Qy 2038 LQECISSAMPKXKPSRLKGNKHSRNMGGIIGEDTLDKIQRDSEHGLSPDSE 2097
Db 2037 LQECISSAMPKXKPSRLKGNKHSRNMGGIIGEDTLDKIQRDSEHGLSPDSE 2096
Qy 2098 NFDWKAIQEGANSIVSSLHO-AAAAACLROASSSDSILSLKSGISLGSPPHLPDQEE 2156
Db 2097 NFDWKAIQEGANSIVSSLHO-AAAAACLROASSSDSILSLKSGISLGSPPHLPDQEE 2156
Qy 2157 KPFTSNKGRPHILKPGKSTLETKKIIESEKIGKGVKYSKLTGKVRNSSEISGOMKQ 2216
Db 2157 KPFTSNKGRPHILKPGKSTLETKKIIESEKIGKGVKYSKLTGKVRNSSEISGOMKQ 2216
Qy 2217 LOANMPSISRGRTMIHIGVRNSSSSTSPVSKGPPKTPASKSPSEGOATTPSRGAKP 2276
Db 2217 LOANMPSISRGRTMIHIGVRNSSSSTSPVSKGPPKTPASKSPSEGOATTPSRGAKP 2276
Qy 2277 SVKSELSPVARQTSQITGGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNISTSPGRNGI 2336
Db 2277 SVKSELSPVARQTSQITGGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNISTSPGRNGI 2336
Qy 2337 SPNKLSQLPRTSSPSTASTKSSGSKMSYTPSGROMSQOONLTQOTGLSKNASSIPRSES 2396
Db 2337 SPNKLSQLPRTSSPSTASTKSSGSKMSYTPSGROMSQOONLTQOTGLSKNASSIPRSES 2396
Qy 2397 ASKGLNOMNANGANKVELSRMSSTKSSGSDRSERPVLRQSTFIKEAPSPTRRLK 2456
Db 2397 ASKGLNOMNANGANKVELSRMSSTKSSGSDRSERPVLRQSTFIKEAPSPTRRLK 2456

Qy	2457	BESASPELSPSRPASPTRSQOATVLSPLDMSLSTHSSVOAGWKLPPNLSPTIE	2516	CC	axin (By similarity).
Db	2457	BESASPELSPSRPASPTRSQOATVLSPLDMSLSTHSSVOAGWKLPPNLSPTIE	2516	CC	-!- ALTERNATIVE PRODUCTS:
Qy	2517	YNDGRPAKRHDIAHSHSESPSRIPINRSQGTWKEHSHSSLPVSTWRTGSSSIIISA	2576	CC	Event=Alternative splicing; Named isoforms=4;
Db	2517	YSDGRPSKRHDIAHSHSESPSRIPVNRAGTWKEHSHSSLPVSTWRTGSSSIIISA	2576	CC	Name=1;
Qy	2577	SSSESSEKAKSEDEKHVNSISGTVKQSNQVSAKGTWRKIKENEFSTNSTSTQVSSGATN	2636	CC	Isoid=Q61315-1; Sequence=Displayed;
Db	2577	SSSESSEKAKSEDEKHVNSVPGQMKENQVPTKGTWRKIKESBITNTVSTQSSGAAS	2636	CC	Name=2;
Qy	2637	GAESKTLIYQMAVAKGTEDVWVRIEDCPINNPRSGRPTGTPPVIDSVSFKANPNIKD	2696	CC	Isoid=Q61315-2; Sequence=VSP_004116;
Db	2637	GAESKTLIYQMAVAKGTEDVWVRIEDCPINNPRSGRPTGTPPVIDSISEKGNPSIKD	2696	CC	Name=3;
Qy	2697	SKNQAKQNVGSGVPMRTVGLNRLNSFIQVDAPQKGTETKPGQNNVPVSEINNESSI	2756	CC	Isoid=Q61315-3; Sequence=VSP_004117;
Db	2697	SKDTQKQSGVSGSGS-PVQTVGLNRLNSFIQVAPSKGTETKAGQSGSPAPVAETGETCM	2755	CC	Name=4;
Qy	2757	VERTPPSSSSSKHSSPSGTVAAARVTFPNVNPSPKSSADTSARPSQIPTPVNNTKKR	2816	CC	Isoid=Q61315-4; Sequence=VSP_004116, VSP_004117;
Db	2756	AERTPPSSSSSKHSSPSGTVAAARVTFPNVNPSPKSSADTSARPSQIPTPVGSSTKKR	2815	CC	TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	lung, brain, stomach, intestine, testis and ovary.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	-!- PTM: Phosphorylated by GSK3B (By similarity).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	-!- SIMILARITY: Contains 7 ARM repeats.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	-----
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	the European Bioinformatics Institute. There are no restrictions on its
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	use by non-profit institutions as long as its content is in no way
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	modified and this statement is not removed. Usage by and for commercial
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	or send an email to license@isb-sib.ch).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	-----
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	EMBL; M88127; AAB59632.1; -;
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	EMBL; U02937; AAA03443.1; -;
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	PIR; I49505; I49505.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	HSP; Q02248; 3BCT.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	MGI; M8039; Apc.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	GO; GO:0005737; C:cytoplasm; IDA.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	GO; GO:0005634; C:nucleus; IDA.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	GO; GO:0008013; P:beta-catenin binding; IDA.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	GO; GO:0009953; P:axis specification; IMP.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	GO; GO:0016053; P:Wnt receptor signaling pathway; IDA.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	InterPro; IPR008938; ARM.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	InterPro; IPR000225; Armadillo.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	Pfam; PF00514; Armadillo_seg; 4.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	SMART; SM00185; ARM; 5.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	PROSITE; PS0176; ARM_REPEAT; 1.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	Wnt signaling pathway; Anti-oncogene; Phosphorylation;
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	Alternative splicing; Repeat; Coiled coil.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	COILED COIL (POTENTIAL).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	COILED COIL (POTENTIAL).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	LEU-RICH.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	ARM 1.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	ARM 2.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	ARM 3.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	ARM 4.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	ARM 5.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	ARM 6.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	ARM 7.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	SER-RICH.
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	ASP/GLU-RICH (ACIDIC).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	ASP/GLU-RICH (ACIDIC).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	HIGHLY CHARGED.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	Missing (in isoform 2 and isoform 4).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	/FTid=VSP_004116.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	Missing (in isoform 3 and isoform 4).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	/FTid=VSP_004117.
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	T -> A (IN STRAIN CAST/EI).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	V -> I (IN STRAIN CAST/EI).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	Y -> F (IN STRAIN CAST/EI).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	A -> T (IN STRAIN CAST/EI).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	A -> S (IN STRAIN CAST/EI).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	H -> A (IN STRAIN CAST/EI).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	G -> Q (IN STRAIN CAST/EI).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	T -> S (IN STRAIN CAST/EI).
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	T -> S (IN STRAIN CAST/EI).
Qy	2817	DSKTDSTESSGTSQSPKSHSGSYLVTSV	2843	CC	SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;
Db	2816	DSKTDSTESSGAQSPKSHSGSYLVTSV	2842	CC	Query Match 90.5%; Score 13178; DB 1; Length 2845;

Best Local Similarity 90.2%; Pred. No. 0; Matches 2572; Conservative 111; Mismatches 155; Indels 12; Gaps 10;

QY 1 MAAASYDQLLKQVEALKXENSNLRQLEDNSNHLTKLETEASNMEVLKQLQGSIEDRAM 60
 Db 1 MAAASYDQLLKQVEALKXENSNLRQLEDNSNHLTKLETEASNMEVLKQLQGSIEDRAM 60

QY 61 ASSGOIDLLELRELKELNLUSSNPPGYKLSKMSLRSGREGSVSRSGCECPVPMGSPPR 120
 Db 61 -TSGOIDLLELRELKELNLUSSNPPGYKLSKMSLRSGREGSVSRSGCECPVPMGSPPR 118

QY 121 RGFVNGSRESTGYLBELEKERSLLADLDKEKEKDYWYAOQLNLTKEIDSLPTENPSL 180
 Db 119 RGFVNGSRESTGYLBELEKERSLLADLDKEKEKDYWYAOQLNLTKEIDSLPTENPSL 178

QY 181 QTDWTRQLEYBARQIRVAMEBQJTCODMEKRAQRRIARQQEKOILRIQLLOSOAT 240
 Db 179 QTDWTRQLEYBARQIRVAMEBQJTCODMEKRAQRRIARQQEKOILRIQLLOSOAA 238

QY 241 EAERSQNKHETGSHDAERQEGQVGBINNAISNGOGSTTRMDHETASVLSSTHSA 300
 Db 239 EAERSQSHDAERQEGQVGBINNAISNGOGSTTRMDHETASVLSSTHSA 298

QY 301 PRRLTSHLGTKEVYVLSLMLGTHDKDDMGRTLLAMSSQSDSCISMPOSGCLPLLIQLL 360
 Db 299 PRRLTSHLGTKEVYVLSLMLGTHDKDDMGRTLLAMSSQSDSCISMPOSGCLPLLIQLL 358

QY 361 HGNDKDSVLLGNSRGSKEARAPASAAALHNIHSDPDDKRGREIRVHLHLOIRAYCETC 420
 Db 359 HGNDKDSVLLGNSRGSKEARAPASAAALHNIHSDPDDKRGREIRVHLHLOIRAYCETC 418

QY 421 WEWQBAEPFGDQDKNPMPAPVEHQICPAVCVLMKLSFDEEHRHANNELGGQIAIELQ 480
 Db 419 WEWQBAEPFGDQDKNPMPAPVEHQICPAVCVLMKLSFDEEHRHANNELGGQIAIELQ 478

QY 481 VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSEEDL 540
 Db 479 VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSEEDL 538

QY 541 QQVIASVLRLNLSWRADVNSKTLREVGSKVAMECALEVKKESTLKSVLALWNLASHCT 600
 Db 539 QQVIASVLRLNLSWRADVNSKTLREVGSKVAMECALEVKKESTLKSVLALWNLASHCT 598

QY 601 ENKADICAVDGALAPLVGLTVRSQNTNLATIESGGILTRNVSSLATNEDHRLRENN 660
 Db 599 ENKADICAVDGALAPLVGLTVRSQNTNLATIESGGILTRNVSSLATNEDHRLRENN 658

QY 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKOEAALWDMGAVSMKLNLIHSHKMIAM 720
 Db 659 CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKOEAALWDMGAVSMKLNLIHSHKMIAM 718

QY 721 GSAALRLNLMANRPAPKYKDNANIMPGSSLSPLHVRKQKALEAEALDAQHLSETFDNIDLS 780
 Db 719 GSAALRLNLMANRPAPKYKDNANIMPGSSLSPLHVRKQKALEAEALDAQHLSETFDNIDLS 778

QY 781 PKASHRSKORHKQSLYGDYVDTNRHNDNRDNFTGNMTVLSPLNTTVLPSSSSSRGS 840
 Db 779 PKASHRSKORHKQSLYGDYVDTNRHNDNRDNFTGNMTVLSPLNTTVLPSSSSSRGS 838

QY 841 LDSRSEKDRSLERERGLGNYPHATENPGTSSKRGICQISTTAAQAKWEEVSAIHTS 900
 Db 839 LDSRSEKDRSLERERGLGNYPHATENPGTSSKRGICQISTTAAQAKWEEVSAIHTS 898

QY 901 QEDRSSTGTTLHCVTDERNALRSSAAHTNSNTYNTFTKSNRNTCSMPYAKLEYKRSS 960
 Db 899 QEDRSSTGTTLHCVTDERNALRSSAAHTNSNTYNTFTKSNRNTCSMPYAKLEYKRSS 958

QY 961 NDSLNSVSSDGYGRGQMKPSIESYEDDESKEFCSCYQYPADLAHKHSHANHDDNDGE 1020
 Db 959 NDSLNSVSSDGYGRGQMKPSIESYEDDESKEFCSCYQYPADLAHKHSHANHDDNDGE 1018

QY 1021 LDTPINYSLKYSDEQLNSGRQSPQNERWAPKHIIIDEIKQNEQROARSQNTSYPVYSE 1080
 Db 1019 LDTPINYSLKYSDEQLNSGRQSPQNERWAPKHIIIDEIKQNEQROARSQNTSYPVYSE 1078

QY 1081 STDDKHLKFPQHPHQECVSPYSRGANGSETNRRVGSNHNQINQVOSLQCOEDDYEDDKP 1140
 Db 1079 NTDDKHLKFPQHPHQECVSPYSRGANGSETNRRVGSNHNQINQVOSLQCOEDDYEDDKP 1138

QY 1141 TNSERYSEEOH-EHEERTNYSIKYNEBKRVHDQIDYSLKYATDIPSSQKQSFPSFK 1199
 Db 1139 TNSERYSEEOH-EHEERTNYSIKYNEBKRVHDQIDYSLKYATDIPSSQKQSFPSFK 1198

QY 1200 SSGQSKTEHMSSEBENTSPSSNAKRONQLHPSSAQSRSGQPOKAATCKVSSINQETI 1259
 Db 1199 NSQAQSTKPEHLSPSSSENTAVPPSNKRNQLPSSAQ-RNGOTQKGTTCVKVSIQETI 1257

QY 1260 QTYCVEDTPTCFGRCSLSLSAEBDEIGCNOQTTOEADSNANTLOIAIEKIKITRSAEDP 1319
 Db 1258 QTYCVEDTPTCFGRCSLSLSAEBDEIGCNOQTTOEADSNANTLOIAIEKIKITRSAEDP 1317

QY 1320 VSEVPAVQSHPRKTSRLOGSSLSSESARH-KAVEFSSGAKSPSKSGAOTPKSPBHYVQ 1378
 Db 1318 ATEVPAVQSHPRKTSRLOGSSLSSESARH-KAVEFSSGAKSPSKSGAOTPKSPBHYVQ 1377

QY 1379 ETPLMFSRCTSVSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPSDQOTMPPSRSKT 1438
 Db 1378 ETPLMFSRCTSVSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPSDQOTMPPSRSKT 1437

QY 1439 PPPPQTAQTKRVKPKAPTAERKRSKQKAAVNAVQVQVLPADTLLHFAETESTPD 1498
 Db 1438 PPPPQTAQTKRVKPKAPTAERKRSKQKAAVNAVQVQVLPADTLLHFAETESTPD 1497

QY 1499 GFSCSSLSALSLEDEFFIQKVELRIMPPVQENDNGNETESEOPEESNENQKEAKTID 1558
 Db 1498 GFSCSSLSALSLEDEFFIQKVELRIMPPVQENDNGNETESEOPEESNENQKEAKTID 1556

QY 1559 SEKDLDDSDDDDIIELECIISAMPTKSRKAKPAQATASKLPPPVARKPSOLPVYKLL 1618
 Db 1557 SEKDLDDSDDDDIIELECIISAMPTKSRKAKPAQATASKLPPPVARKPSOLPVYKLL 1616

QY 1619 PSQNRLOPOKHVFTPGDDMPRVYCVGPTPIFSTATISDLTIESPPNELAAGEVRGG 1678
 Db 1617 PAQNRLOPOKHVFTPGDDMPRVYCVGPTPIFSTATISDLTIESPPNELAAGEVRGG 1676

QY 1679 AQSGFEKRDITPTEGRSTDEAOGGKTSVTPPELDDNKAEGDILAEICINSAMPKGS 1738
 Db 1677 IQSGFEKRDITPTEGRSTDEAOGGKTSVTPPELDDNKAEGDILAEICINSAMPKGS 1736

QY 1739 KPRVRKIMDOVOOASASAPNKNQIDGKKKPTSPVKPIPONTYRTRVRKNADSKN 1798
 Db 1737 KPRVRKIMDOVOOASASAPNKNQIDGKKKPTSPVKPIPONTYRTRVRKNADSKN 1796

QY 1799 LNAERVFSNMDKSKQNLKNNKDFNDKLPNNEDRVGSGFAFDSPHYTPTEGTPYCFSR 1858
 Db 1797 VNTTEFSDNKDSKPSLQTNAKAFNEKLPNNEDRVGSGFAFDSPHYTPTEGTPYCFSR 1856

QY 1859 NDSLSDLDLDDDDVDLSREKAEURKAKENKESPAKVTSHTELTSNOQSKANTQAIKQPI 1918
 Db 1857 NDSLSDLDLDDDDVDLSREKAEURKAKENKESPAKVTSHTELTSNOQSKANTQAIKQPI 1916

QY 1919 NRQOPKILQKQSTFPQSSKDIIDRGAATDEKLFONFAIENTPVCFSHNSLSLSLSDIDR 1978
 Db 1917 NRAQSPVLOKQSTFPQSSKDIIDRGAATDEKLFONFAIENTPVCFSHNSLSLSLSDIDR 1976

QY 1979 -NNKNENPIKETEPDPSQCEPSKQASGYAPKGFHVEDTTPVCFSRNSSLSSLSIDSEDD 2037
 Db 1977 -NNKNENPIKETEPDPSQCEPSKQASGYAPKGFHVEDTTPVCFSRNSSLSSLSIDSEDD 2036

QY 2038 LLOECISSAMPKKKPSRLKGNNEKHSPRNMGGILGEDLTLDLQIQRDPSHGLSPDSE 2097
 Db 2037 LLOECISSAMPKKKPSRLKGNNEKHSPRNMGGILGEDLTLDLQIQRDPSHGLSPDSE 2096

QY 2098 NFDWKAIQEGANSIVSSLHQ-AAAAACLSRQASDSDSILSKSGISLSPHLLTDPQSE 2156
 Db 2097 NFDWKAIQEGANSIVSSLHQ-AAAAACLSRQASDSDSILSKSGISLSPHLLTDPQSE 2156

RA Chen F., Rebay I.;
 RT "split ends, a new component of the Drosophila EGF receptor pathway,
 RL regulates development of midline glial cells.";
 RN Curr. Biol. 10:943-946(2000).
 [9]
 RP FUNCTION ON WG PATHWAY.
 RX MEDLINE=22668876; PubMed=12783785;
 RA Lin H.V., Doroquez D.B., Cho S., Chen F., Rebay I., Cadigan K.M.;
 RT "Split ends is a tissue/promoter specific regulator of Wingless
 RL signaling.";
 RL Development 130:3125-3135(2003).
 CC -!- FUNCTION: Probable corepressor protein, which regulates different
 CC key pathways such as the EGF receptor and Wg pathways. Involved in
 CC neuronal cell fate, survival and axon guidance, cell cycle
 CC regulation and repression of head identity in the embryonic trunk.
 CC May act with the Hox gene Deformed and the EGF receptor signaling
 CC pathway. Positive regulator of the Wg pathway in larval tissues
 CC but not in embryonic tissues. May act as a transcriptional
 CC corepressor protein, which repress transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative promoter;
 CC Comment=2 isoforms, 1 (shown here) and 2, are produced by use of
 CC alternative promoters;
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=1;
 CC IsoId=Q8SX83-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8SX83-2; Sequence=VSP_008565, VSP_008566;
 CC Note=No experimental confirmation available;
 CC Name=3; Synonyms=SpnL;
 CC IsoId=Q8SX83-3; Sequence=VSP_008567;
 CC Note=Produced by alternative splicing of isoform 1;
 CC Name=4; Synonyms=SpnS;
 CC IsoId=Q8SX83-4; Sequence=VSP_008565, VSP_008566, VSP_008567;
 CC Note=Produced by alternative splicing of isoform 2;
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization
 CC in stage 3 embryos, and in blastoderm cells, including pole cells.
 CC Expressed throughout the rest of embryogenesis. Later, it is
 CC expressed at higher level in epidermal cells and CNS.
 CC -!- DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and
 CC zygotically.
 CC -!- SIMILARITY: Belongs to the Spn family.
 CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
 CC -!- SIMILARITY: Contains 1 SPOC domain.
 CC -!- CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF188205; AAF13218.1; -;
 CC EMBL; AF184612; AAF26299.1; -;
 CC EMBL; AF221715; AAF34661.1; ALT_INIT.
 CC EMBL; AE003590; AAF51534.2; -;
 CC EMBL; AE003590; AAF51535.2; -;
 CC EMBL; AE003590; AAN10511.1; -;
 CC EMBL; AY094788; AAM11141.1; ALT_SEQ.
 CC HSPG; P09651; IHA1.
 CC FlyBase; FBgn0016977; spen.
 CC DR GO; GO:0007411; P:axon guidance; IMP.
 CC DR GO; GO:0008347; P:glia cell migration; IMP.
 CC DR InterPro; IPR000504; RNA_rec_mot.
 CC DR Pfam; PF00076; rrm; 3.
 CC DR SMART; SM00360; RRM; 3.
 CC DR PROSITE; PS0102; RRM; 3.
 CC DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.

DR PROSITE; PS00917; SPOC; 1.
 KW Transcription regulation; Repressor; Developmental protein;
 KW Nuclear protein; Repeat; RNA-binding; Coiled coil;
 KW Alternative promoter usage; Alternative splicing.
 FT DOMAIN 554 632 RNA-BINDING (RRM) 1.
 FT DOMAIN 656 730 RNA-BINDING (RRM) 2.
 FT DOMAIN 734 806 RNA-BINDING (RRM) 3.
 Query Match 3.3%; Score 474; DB 1; Length 5560;
 Best Local Similarity 18.5%; Pred. No. 8.3e-10;
 Matches 621; Conservative 479; Mismatches 1219; Indels 1034; Gaps 147;
 QY 2 AASYDQLLKQVE--ALKWENLNLEQLEDNSNHLTKLETEASNKMKVILKQLOGSIEDE 58
 DB 1786 SSLSLSTTASVTASISISSSTNGNSLTSAAHVQKFCQSTFVEEHTKSGTSTQS 1845
 QY 59 AMASSQIDLLERKELNLDSSNFPQKLRSLKMSLRSGRSGVSSRSGECSVPWGSF 118
 DB 1846 SSSSSKKI-----SSTHD--KLHSHNNRSESDKKIKKSDKNASSDKRKNSS 1891
 QY 119 PRGFVNGRSTGYLEE-----LEKERSLLADLDKEEK 153
 DB 1892 TSQ-----SKSATPRIEDDSSEADDTADKAENQHEKKEKREKDLKQVREK 1947
 QY 154 EKDWYIAQLNLTFRIDSLPTENFSLQDMTRQLEYEARQIRVAMEBQLGTCQDWEKR 213
 DB 1948 DRK--AQEEREK-----EDRAKEEKEEREKKAQEDREKKEEREELREKEQ 1996
 QY 214 -----AQRIARIQIEKD-----IIRIQLLOSQATEARRSQNH-ETGSHDAER 259
 DB 1997 DKEQKEIREKDLREKEQERDRNREKELDKDLREKREKEQREKELHREKQDREH 2056
 QY 260 QNEGQGVGEINMATSGNGOGSTTRMDHETASVLSSSTSHAPRLTSLHGTQVEM---VY 316
 DB 2057 REKEQRRAMDVEQERG-----GRM-----RELSSYQSKMDIAGEAS 2095
 QY 317 SLISMLGTHDKD-----DMGRTLLAMS-----SSQDS 343
 DB 2096 SLTAIDCQHNKENAMDTIAQTGASPTSDNTPKERSKLSRNPVRLHKKRLSSQES 2155
 QY 344 CISMRSGCLPLLIQLHNGD-----KDSVLLGNRSGKEAR-----380
 DB 2156 NBSAGGGSGGSGSHQIHEDYVYKIRIMENSONISVHSSNQRLNDRRDSKEHSSSFKE 2215
 QY 381 -----ARASAAALNIIHSPDDKGRREIR-----VLHLEQIRAYCE 418
 DB 2216 KNSSSHISRHPCGGSSASSSKHHRRDKHQKGSASSIETNWSIEVVDPISQTKHNLN 2275
 QY 419 TCWEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMMELGQLQIAEL 478
 DB 2276 TSEELQSHQPKREKER-----EHFSGHANSSSSRHKSXRDHHRHREK--KRHSVAES 2326
 QY 479 LQVDCMYGLT-NDHYSITLRRYAGMALTNLPDGVANKATLCMKGCMRALVAQLKSES 537
 DB 2327 TWYDEHTTQQHNPHRIS-----AAGSGAGELSSAATNTS-----2363
 QY 538 EDLQVVIASVLRNLWRADVNGKTLREVGSKVYKALMECALEVKESTLKVLSALNLSA 597
 DB 2364 -----SGKLHHQHRRSVERKSSR---GS-----DEGHHSSSKSLAKMLMSS 2404
 QY 598 HCTENKADICAVDGAFLVGLITYESQNTLAIIESGGGILRNVSLSLATNEDHRIQLR 657
 DB 2405 -----ADSDDTDDA-----2413
 QY 658 ENNCLQTLQHLKSHSLTIVSNACGTLWNLSARNPKDQBALMDMGAVSMLKNL-----710
 DB 2414 -----SKKHSIFDIPDDCPNV-----SMYDKVKARCKNQKQAEK 2450
 QY 711 -HSKHKHMTAMGSAALRNLM--ANRPAYKDNAMSPGSSLPPLHVR-----755
 DB 2451 KIKAKTSQLKQSRKAKKSTSYDGSDDTEFEDRQHRNNGSS--SFGRYPLGLSSDDDD 2508
 QY 756 ---KQKALEAELDAQHLSSETFDNIDNLSPKASHRSKQRE-----K 792

QY 2630 VSSGATNGAESKTLIYQWAPVSKTEDVWVRIEDC--PINNPRSGRGT-GNTTTPVDSV 2686
 DB 4440 ISSVAPSAEDQRP--RLILTINKTQPSIKNISEMEQTIQQQQQQQSEVITNDPIGGDN 4496
 QY 2687 SEKANPNIKSKDNOAKON-----VNGSVPMRTVGLNRLNSFIQVDAPD-QK 2734
 DB 4497 SESC-NTRKSRRLQEKEDRSTVDIIEVVVNTNTPTGT-----GPHLPK 4540
 QY 2735 GTEIKP--GQN-----NPV-----PVSETNESSIVERTPPSSSSSSK 2769
 DB 4541 GAOPTPRSGRGAQAKTDAVOIINAVGRPRSRDKRTIGQTANLIEEVTASNTVAAS 4600
 QY 2770 HSSPSGTVAR-----VTPNYPNPSRKSSADSTSRPSQIP 2806
 DB 4601 HLAPEGAGVESHVPQLDAKEVEPVSVVTPIS-TPAPVSAAPTVVPVPAWP 4652
 RESULT 6
 MINT_HUMAN
 ID MINT_HUMAN STANDARD; PRT; 3664 AA.
 AC Q9GT58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
 GN MINT OR SHARP OR KIA0929.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
 RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
 RP RAR AND MTA11.
 RC TISSUE=Liver, and Pituitary;
 RX MEDLINE=21231190; PubMed=11331609;
 RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
 RA Hon M., Evans R.M.;
 RT "Sharp, an inducible cofactor that integrates nuclear receptor
 RT repression and activation.";
 RL Genes Dev. 15:1140-1151(2001).
 [2]
 RP SEQUENCE FROM N.A.
 RA Bird C.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 294-3664 FROM N.A.
 RA Rhodes S., Huckle E.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
 RC TISSUE=Embryo, and Teratocarcinoma;
 RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE OF 2002-3664 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kozani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:63-70(1999).
 [6]
 RP INTERACTION WITH PPARD.
 RX MEDLINE=21874127; PubMed=11867749;

RA Shi Y., Hon M., Evans R.M.;
 RT "The peroxisome proliferator-activated receptor delta, an integrator
 RT of transcriptional repression and nuclear receptor signaling.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
 [7]
 RP FUNCTION, AND INTERACTION WITH RBPSUH.
 RX MEDLINE=2261914; PubMed=12374742;
 RA Oswald F., Koscezka U., Astrahantseff K., Bourtelee S., Dillinger K.,
 RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Lipray S.,
 RA Schmid R.M.;
 RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
 RT pathway.";
 RL EMBO J. 21:5417-5426(2002).
 [8]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
 RX MEDLINE=22777836; PubMed=12897056;
 RA Ariyoshi M., Schwabe J.W.R.;
 RT "A conserved structural motif reveals the essential transcriptional
 RT repression function of Spen proteins and their role in developmental
 RT signaling.";
 RL Genes Dev. 17:1909-1920(2003).
 CC -!- FUNCTION: Essential corepressor protein, which probably regulates
 CC different key pathways such as the Notch pathway. Negative
 CC regulator of the Notch pathway via its interaction with RBPSUH,
 CC which prevents the association between NOTCH1 and RBPSUH, and
 CC therefore suppresses the transcriptional activity of Notch
 CC signaling. Blocks the differentiation of precursor B cells into
 CC marginal zone B cells. Probably represses transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins. May bind both to DNA and RNA.
 CC -!- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
 CC NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTA11. Interacts with
 CC RBPSUH; this interaction may prevent the interaction between
 CC RBPSUH and NOTCH1. Interacts with the nuclear receptors RAR and
 CC PPARD. Interacts with RAR in absence of ligand. Bind to the
 CC steroid receptor RNA coactivator SRA.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
 CC -!- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
 CC spleen and thymus. Expressed at intermediate level in kidney,
 CC liver, mammary gland and skin.
 CC -!- INDUCTION: By hormone 17-beta-estradiol (E2).
 CC -!- DOMAIN: The RID domain mediates the interaction with nuclear
 CC receptors (By similarity).
 CC -!- DOMAIN: The SPOC domain, which mediates the interaction with
 CC NCOR2, is essential for the repressive activity.
 CC -!- SIMILARITY: Belongs to the Spen family.
 CC -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
 CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
 CC -!- SIMILARITY: Contains 1 SPOC domain.
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 CC EMBL; AF356524; AAK52750.1; -;
 DR EMBL; AL034555; CAB85442.1; ALT_SEQ.
 DR EMBL; AL034555; CAB85444.1; ALT_SEQ.
 DR EMBL; AL450998; -; NOT_ANNOTATED_CDS.
 DR EMBL; AL096858; CAB51072.1; ALT_INIT.
 DR EMBL; AK000882; BAA91405.1; ALT_INIT.
 DR EMBL; AK022949; BAB14324.1; ALT_INIT.
 DR EMBL; AB023146; BAA76773.1; -;
 DR InterPro; IPR000504; RNA_rec_mot.
 DR PDB; 1OW1; 19-AUG-03.
 DR Pfam; PF00076; rrm; 4.
 DR SMART; SM00360; RRM; 4.
 DR PROSITE; PS50102; RRM; 4.

DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 DR PROSITE; PS00917; SPOC; 1.
 KW Transcription regulation; Repressor; Nuclear protein; DNA-binding;
 KW Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
 FT DOMAIN 1 573
 FT DOMAIN 6 81
 FT DOMAIN 335 415
 FT DOMAIN 438 513
 FT DOMAIN 517 589
 FT DOMAIN 688 715
 FT DOMAIN 977 1004
 FT DOMAIN 1170 1191
 FT DOMAIN 1408 1428
 FT DOMAIN 1496 1529
 FT DOMAIN 1592 1612
 FT DOMAIN 1928 1944
 FT DOMAIN 2201 2707
 FT DOMAIN 3498 3664
 FT DOMAIN 2130 2464
 FT DOMAIN 2709 2870
 FT DOMAIN 125 277
 FT DOMAIN 240 325
 FT DOMAIN 616 810
 FT DOMAIN 624 697
 FT DOMAIN 2428 2520
 FT DOMAIN 3220 3482
 FT VARIANT 970 970
 FT VARIANT 1091 1091
 FT VARIANT 2360 2360
 FT VARIANT 956 956
 FT CONFLICT 3664 AA; 402245 MW; 5228C58533E5B27B CRC64;
 FT SEQUENCE

Query Match 3.1%; Score 452; DB 1; Length 3664;
 Best Local Similarity 19.3%; Pred. No. 3.3e-09;
 Matches 490; Conservative 341; Mismatches 969; Indels 742; Gaps 115;

QY 727 RNLMAIRPAK-YKDNIMSPGSLPSLHVKKQKALEDAQLHSETFDNIDNLSPK-A 783
 DB 226 REYGRPRPENYCHSRSPHSS-----QSRNQSPQLA 259
 QY 784 SHRSQRHQSLQYGVYFDNTHDNRDNFNFTGNVLPVLYNTVLPSSSSSSGSLDS 843
 DB 260 SQASRPTRPSGSG-----SRSSSSSDS188-----SSSTSSSDSDS 297
 QY 844 SRSEKD-----RSLERE---RGIGLGNVHPATENPGTSSKRGLOI 880
 DB 298 SSSSSDSDSPARSVQSAAPAPTSQLLSSLEKDEPKRSGFKVQNLVRSYTDLSLQGLFH 357
 QY 881 -----STTAAQIAKYMEEVSAT--HTSQEDRSSGSTTT-----LHCYVD 917
 DB 358 EFKKFGKVTQVIGHTSEERYGLVFRQDQEKALTASKGLFGQMIEVTAWIGPTE 417
 QY 918 ERNALR-----RSSAHT-----HSNTYNFTKS-----ENNRCTCMP- 950
 DB 418 SENEFPALDERIDEHPKATRTLFIGNLEKTTVYHDLNRFORFGEIVDIDIKYNGVPQ 477
 QY 951 YAKLEY-----KRSSNDSLSVSSSDGYGRGQMKPS-----IESYEDDESK--- 993
 DB 478 YAFLOQCDIASVCKAIKQWDGEVIGNRLKLGFGK---SWPTNCVWLGLSSNVSDQYLT 534
 QY 994 --FCSTG-----QYPADLAHKIHSANHMDNDGELDTPNYS-L 1029
 DB 535 RHFCRYGPVVKVYFDRKLGKVALVYNEIEY-AQAAVKETKRGKIGNKIKVDFAKRESOL 593
 QY 1030 KYSDQELNSGRQSPQNERWARPKHIEDEIKQSEORQS--RNQSTTY-----PVYTE 1080
 DB 594 AFVHCVEKSGQDIRDFYEMLAEE-----RREERASYDYNQRTYVESVRTPTYPE 644
 QY 1081 -STDDKHLKQPHFGQBCVSP--YASRGANGSENVNRSNGINQNVQSQCLEDDEYD 1137

DB 645 DSRDYPARGREFYSEWETVQGYESRYDDPREVDYRNDPEQDI-----REYSYRQ 699
 QY 1138 DKPTNYSERYSEEBEQHEERPTNYSIKYNEKRHVDPID--YSLKYATDIPS-SOKQS 1194
 DB 700 RERERERFESDRDRDHERRP---IERQSPVHLRARPQSPGASPSQAERLPSDSERRL 755
 QY 1195 PSFKSSSGGSSK-----TEHSSSSSENTSTPSSNAKQNLHPSAQ 1237
 DB 756 YSRSDRSGCSSLSPRYEKLDRSLRYTKNEKTDFPERVERERIRKEKVE 815
 QY 1238 --SRSGOPKAATCKVSSINQETIQTVCVEDTPICFSRCSLSSLSAABEIGCNOTTOE 1295
 DB 816 KDKTDKQKRKGKVPSSQSSSETDOENEREOSPEKPRSCNKLREKADKEGIAKNR--- 871
 QY 1296 ADSANTLOIAIEKIKIGTRSAEDPVSEVPVAVSOHPTKSSRLQ-----GSLSS 1344
 DB 872 LELMPCVVLTRVKEKGV-DHTPTVEKLAKLNDTVKSSALDQKLQVSGTEPAKSDLSK 931
 QY 1345 -ESARHKAIVE---PSSGAKSPKSGAOTPKSPPEHYVQETPLMFSRCTSVSSLSDSPESR 1400
 DB 932 LESVRMKVPKEGLSSHVVEKEGLKAR--KHLKPEQADGVSAVDLEKLEARKERF 988
 QY 1401 IASSVQSEPCSGMVSGIISPSDLPDSPQTMPPSR---SKTPPPPTQATQKREVPNKA 1457
 DB 989 ADSNLKAEK-----QKPEVKKSP--EMEDARVLSKKQPDVSSREVILLREGEAERK 1038
 QY 1458 PTAB--KRESGPKQAAVNAVQVLPDADTLHPATESTPDGFCSSSLSALSILDEP 1514
 DB 1039 PVKELIKRES-----KKIK-----LDRLNTVASPKDQELASISVSGSGSRP 1080
 QY 1515 FIQKDELIRMPVQENDNGNTESEQKPSNENQEKAEKTIIDSKDL-----LDDSD 1569
 DB 1081 --SSDLQARL-----GELAGSEVENQEVQSKKPIPSKPOLKQLQVLDDQGP 1124
 QY 1570 DDTEILEE-CII--SAMPTKSSRKAKKPAOTASKLPPVARKPSQLPVYKLLSQRELQP 1626
 DB 1125 EREDVRKNYCSLRDETPERKSGQKSHSVNTEKIGIDHTQS---YKQMEQSRKKQ 1180
 QY 1627 QKHVSF-----TPGDDMPRVYCVETPINFSTATSLDLTIESPPNELAAGEV--- 1675
 DB 1181 QWEMEIAKSEKFGSPKQVDE---YEERSLVHVGKPPQDVTDSPSKKRMHDVDFDI 1237
 QY 1676 -----RGAQSGEPEKDDTPTTEGRST-----DEAQGKTSSVTI---PBLDDN 1716
 DB 1238 CTKEERNYRSRQISEDSER---TGGSPSVRHGSHFEDDPICSPRLLSVKGSPKVDK 1293
 QY 1717 KAEGBDILASINSAMPKGSHKPFVKIMQVQ-QASASSAPNKNQLDGKKK--- 1771
 DB 1294 VLPYSNI---TVREESLKFPYDSSREEQVADMAKIKLVNLSDEDLNRWDSQMKQDAGR 1350
 QY 1772 ---PTSPVKPIQNTYETVRKNADSKNNNAERVSFSDNKKDKKKQNLKNNNS----- 1820
 DB 1351 FVVSFPNSIHK-----RDSLKR--SVRDLEPGEVPSDSDEDEGH--KSHSPRASAL 1398
 QY 1821 -----KDRNDKLPNNEDVRG-----SFAFDSPHVHTPIEGTFCFSNDSL 1862
 DB 1399 YESRLSFLDRDEDKLRERDERLSSLRNKYFISALDK---TTPDTKALLERAKSL 1454
 QY 1863 SS-----LDFDDDDVDL-----SREKAE-----LRKAKENKSEAKV---TS 1896
 DB 1455 SSSREENWSFLWDSRFANFRNNKKEKVDSPAPRIPSWYMKKKKIRTDSGKDDKXED 1514
 QY 1897 HTELTSNQOQANKTQ---AIKQPINRGQPKILQKQSTFPOS---SKDIPDRGAATDE 1949
 DB 1515 HKEREQERQLFASRFLHSSIFEDSKRLOHLERKEEDSDFISGRYKQTSNGANSTTD 1574
 QY 1950 KLOQFAIENTFVCFSHNSLSSLSLSDIDQENNNKENEPKETEPPDQSGEPKQASGYAP 2009
 DB 1575 SIQE-----PVLVLF--SRMELTRMQKEKEKQKPEKEQEDTENHPKTPES---AP 1624
 QY 2010 KSHVBE-DTPVCFSRNSSLSSLSLSDIDDLQBCISSAMPKKKPKSRLLKGDNEKHSFRNM 2068
 DB 1625 ENKDSLUKTPPSVG--PPSVTVTLESAPSALE-----KTTGDKTVEAP--- 1666

QY 2069 GGIIGEDLTLDLKDILQDPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAAAACLSRQA 2128
DB : : : : :
QY 1667 --LVTEETVE-----PATVSEAPAPAPAEQLEQVD-----LSPGA 1706
DB : : : : :
QY 2129 SSDSDIISLKSGLSGSPFHLTPDQEEKPTSNKGRILKPKGKSTLETKKIESKGI 2188
DB : : : : :
QY 1707 DPDKAAM-MPAGVEEGSGGQPPYLDKAPT-----PGASFQAESNVDP-- 1752
DB : : : : :
QY 2199 KGGKKVYKSLITGKVRNSEISGQMKQPLQANMPSISGRGRMHIHPGVRNNSSTSVSK 2248
DB : : : : :
QY 1753 -----PDSQPLSK 1761
DB : : : : :
QY 2249 KGPPPLKTPASKSPBEGOTATTSRPAKPSVKSELSPVARQTSQ----- 2291
DB : : : : :
QY 1762 -----PAQKSEANE-----PRAEKDDATADAEPAQVAAAPESQPPASEDLEVP 1809
DB : : : : :
QY 2292 -IGSSKAPSGSGRDSPTSPRA-----QQPLSRPIQSPGRNISPGRNGISPPNKLSQL 2345
DB : : : : :
QY 1810 PVAARDKPKNS-KRSKTPVQAAAVSIVEKPVTRKSERIDREKLKRSNPRGEAQKLEL 1868
DB : : : : :
QY 2346 PRTSSPFASTKSSGSGGWSYTSQGRQMSQONLTKQGLSKNASSIPRSASASKGLQNM 2405
DB : : : : :
QY 1869 KWEAEKITRTAKNSAADLEHPELSPLSR-----RRNRV-----RSVATMGDHE-- 1915
DB : : : : :
QY 2406 NNGANKKVELSRMSSTKSSGSGSDRSRPPVLRQSTFIKEAPSPTRLRRKLEBSAS-- 2461
DB : : : : :
QY 1916 NRSPVKEPVEQPRVTRKELRELQEAAPVPTPRG-----RPPKTRRRADEBENEAKE 1970
DB : : : : :
QY 2462 -FESLSPSRASP-----TRSQATPVLSPSL----- 2488
DB : : : : :
QY 1971 PAETLKPPGWRSPRSKTAAGGPGQKKGKNEPKVDATREATEV-GPOIGVKESME 2029
DB : : : : :
QY 2489 PMSLSTHSSVOA-----GGWRKLPPNLSPTIEVNDGRPA--KRHDIARSHSESPSLPI 2541
DB : : : : :
QY 2030 PKAABEEAGSKQDRKDAGTKNPPETAP-VEVVEKKPAPEKNSKKGESRN-SRLAV 2087
DB : : : : :
QY 2542 NRSQGWKHEKSHSSLPVRSVTWARTGSSSLPSSASSGESSKAKSED-----EKHVNISG 2597
DB : : : : :
QY 2088 DKSASLKNVDAVS---PRGAAOAGERESGVAVSPKESPOKEDGLSSQLKSDPVP 2144
DB : : : : :
QY 2598 TKQKENOVSAKGTWRKTKENEFSTPNSTSTVSSGATNGAESKTLVQ-----MAP-- 2649
DB : : : : :
QY 2145 DKEPEKDVASGP--SPEATQLAKQMELEQAVEHIAKLAEASAAAYKADAPGLAPED 2202
DB : : : : :
QY 2650 -----AVSKTE-----DVMVRIEDCPINNRSRSGSPTGNTTPV----- 2682
DB : : : : :
QY 2203 RDKPAHQASSETLAAAGISLINDISGEPENFAPPVPYGESQTDLPAGQAALQPSSEG 2262
DB : : : : :
QY 2693 -----IDSVSEKANPNK-----DSKDNQAKQNVGNGSVPMRTVGLNLR 2721
DB : : : : :
QY 2263 METDEAVSGILETAATESRPPVNPADPPSAGPTDTKEARGNSSETSHSVP-EAKGSKEV 2321
DB : : : : :
QY 2722 LNSFTQVDAPPQKGTETKPGQN-----NPPVNSE-----TNESIVE--RTP 2761
DB : : : : :
QY 2322 EVTLVRKDKGQKTRSRKENTKVKVAVPESHVPSNQAGBSPAANEGTTVQHSAP 2381
DB : : : : :
QY 2762 FSSSSSSK-HSGSPGTVAARVT--PFNNVPS-----PRKSSABSTARSAPQITPPV 2809
DB : : : : :
QY 2382 QEEKQSEKPHSTPPQOSTCSDLSKIPSTENSSQEISVEERTTKASVPPDLPPPPQ-PAPV 2440
DB : : : : :
QY 2810 NNNTKRDSDTSTESSGTQSP 2831
DB : : : : :
QY 2441 DEEPOARFVHSIIESDPVTTP 2462
DB : : : : :

RESULT 7
MINT MOUSE
ID MINT MOUSE
AC Q62504; Q30TN9; Q99P84; Q9QZM2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIAA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
RP LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC TISSUE=Testis;
RX MEDLINE=99379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel mxi2 binding protein, recognizes and
RT regulates the rat osteocalcin promoter.";
RL Biochemistry 38:10678-10690(1999).
RN [2]
RP SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
RA Ono M., Ohnato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA Watahiki A., Muramatsu M., Hayashizaki Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT J/Su(H) activity.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 318-578 FROM N.A.
RC TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozat F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fzanes C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341(1997).
RN [5]
RP SEQUENCE OF 2598-3644 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
RN [6]
RP TISSUE SPECIFICITY.
RX MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kostezka U., Astrahantseff K., Bourteelle S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway.";
RL EMBL J. 21:5417-5426(2002).
RN [7]
RP FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=22483652; PubMed=12594956;
RA Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT "Regulation of marginal zone B cell development by MINT, a suppressor
RT of Notch/RBP-J signaling pathway.";
RL Immunity 18:301-312(2003).
CC -1- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative

regulator of the Notch pathway via its interaction with RBPSUH, which prevents the association between NOTCH1 and RBPSUH, and therefore suppresses the transactivation activity of Notch signaling. Blocks the differentiation of precursor B cells into marginal zone B cells. Probably represses transcription via the recruitment of large complexes containing histone deacetylase proteins. May bind both to DNA and RNA.

BTAIN1: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and interacts with RAR in absence of ligand. Bind to the steroid receptor RNA coactivator SRA (By similarity). Interacts with MSX2. Interacts with RBPSUH; this interaction may prevent the interaction between RBPSUH and NOTCH1.

BTAIN1: SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.

BTAIN1: ALTERNATIVE PRODUCTS:

BTAIN1: Event=Alternative splicing; Named isoforms=2;

BTAIN1: Name=1;

BTAIN1: IsoId=Q62504-1; Sequence=Displayed;

BTAIN1: Name=2;

BTAIN1: IsoId=Q62504-2; Sequence=VSP_008564;

BTAIN1: Note=No experimental confirmation available;

BTAIN1: TISSUE SPECIFICITY: Highly expressed in testis. Expressed at lower level in brain, lung, spleen, liver and kidney. Weakly expressed in cardiac and skeletal muscles and ovary. In spleen, it is expressed in follicular B-cells, while it is weakly expressed in marginal zone B-cells.

BTAIN1: DOMAIN: The RID domain mediates the interaction with nuclear receptors.

BTAIN1: DOMAIN: The SPOC domain, which mediates the interaction with NCOR2, is essential for the repressive activity (By similarity).

BTAIN1: SIMILARITY: Belongs to the Spen family.

BTAIN1: SIMILARITY: Contains 1 RID (receptor interacting) domain.

BTAIN1: SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.

BTAIN1: SIMILARITY: Contains 1 SPOC domain.

BTAIN1: CAUTION: Ref.4 sequence differs from that shown due to multiple frame shifts and conflicts that create stop codons.

BTAIN1: CAUTION: Ref.5 sequence differs from that shown due to what seems to be the presence of intronic sequence in the cDNA.

BTAIN1: This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

BTAIN1: EMBL; BV726481; -; NOT ANNOTATED CDS.

BTAIN1: EMBL; AF156529; AAD55931.1; ALT_INIT.

BTAIN1: EMBL; AB055980; BAB32786.1; -.

BTAIN1: EMBL; Z78160; CAB01562.1; ALT_SEQ.

BTAIN1: EMBL; AK122402; BAC65684.2; ALT_SEQ.

BTAIN1: MGD; MGI:1891706; Mint.

BTAIN1: InterPro; IPR000504; RNA_rec_mot.

BTAIN1: Pfam; PF00076; rrm; 3.

BTAIN1: SMART; SM00360; RRM; 3.

BTAIN1: PROSITE; PS0102; RRM; 4.

BTAIN1: PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.

BTAIN1: PROSITE; PS50917; SPOC_1.

BTAIN1: Transcription regulation; Repressor; Nuclear protein; DNA-binding; RNA-binding; Repeat; Coiled coil; Alternative splicing; Polymorphism.

BTAIN1: DOMAIN 1 574 DNA-BINDING.

BTAIN1: DOMAIN 6 81 RNA-BINDING (RRM) 1.

BTAIN1: DOMAIN 336 416 RNA-BINDING (RRM) 2.

BTAIN1: DOMAIN 439 514 RNA-BINDING (RRM) 3.

BTAIN1: DOMAIN 518 590 RNA-BINDING (RRM) 4.

BTAIN1: DOMAIN 559 575 COILED COIL (POTENTIAL).

BTAIN1: DOMAIN 822 850 COILED COIL (POTENTIAL).

BTAIN1: DOMAIN 1185 1206 COILED COIL (POTENTIAL).

BTAIN1: DOMAIN 1509 1544 COILED COIL (POTENTIAL).

BTAIN1: DOMAIN 1607 1627 COILED COIL (POTENTIAL).

BTAIN1: DOMAIN 2216 2704 RID.

BTAIN1: DOMAIN 3478 3644 SPOC.

FT	DOMAIN	2138	2462	INTERACTION WITH MSX2.
FT	DOMAIN	2706	2845	INTERACTION WITH RBPSUH.
FT	DOMAIN	125	277	ARG-RICH.
FT	DOMAIN	326	326	SER-RICH.
FT	DOMAIN	648	721	TYR-RICH.
FT	DOMAIN	702	832	ARG-RICH.
FT	DOMAIN	2101	2233	ALA-RICH.
FT	DOMAIN	2377	2518	PRO-RICH.
FT	DOMAIN	2950	3475	PRO-RICH.
FT	VARSPLIC	618	640	Missing (in isoform 2).
FT	VARIANT	348	348	I -> T.
FT	VARIANT	762	762	S -> F.
FT	VARIANT	773	773	S -> F.
FT	VARIANT	933	933	S -> L.
FT	CONFLICT	754	754	R -> G (IN REF. 3).
FT	CONFLICT	1524	1524	D -> A (IN REF. 3).
FT	CONFLICT	1560	1560	H -> Y (IN REF. 3).
FT	CONFLICT	1570	1570	F -> L (IN REF. 3).
FT	CONFLICT	1574	1574	R -> G (IN REF. 3).
FT	CONFLICT	1609	1609	Q -> R (IN REF. 3).
FT	CONFLICT	1659	1659	I -> V (IN REF. 3).
FT	CONFLICT	1669	1669	S -> F (IN REF. 3).
FT	CONFLICT	1705	1705	V -> A (IN REF. 3).
FT	CONFLICT	1815	1815	A -> V (IN REF. 3).
FT	CONFLICT	2097	2097	G -> A (IN REF. 3).
FT	CONFLICT	2201	2202	MISSING (IN REF. 3).
FT	CONFLICT	2322	2322	A -> V (IN REF. 3).
FT	CONFLICT	2385	2385	P -> Q (IN REF. 3).
FT	CONFLICT	2502	2502	R -> K (IN REF. 3).
FT	CONFLICT	2505	2505	E -> K (IN REF. 3).
FT	CONFLICT	2519	2519	D -> N (IN REF. 3).
FT	CONFLICT	2554	2554	T -> S (IN REF. 3).
FT	CONFLICT	2679	2688	LVSTPAGPVN -> VGEHPWARD (IN REF. 3).
FT	CONFLICT	3010	3010	L -> P (IN REF. 3 AND 5).
FT	SEQUENCE	3644	3644	AA; 398750 MW; 9C7EC49A51A7DA4A CRC64;
Query Match		2.9%	Score 428.5; DB 1; Length 3644;	
Best Local Similarity		18.7%	Pred. No. 2.4e-08;	
Matches		522; Conservative	325; Mismatches 903; Indels 1047; Gaps 126;	
QY	727	RNLNMPAK-YKDANIMFGSLPSLHVKKQKALEAELDAHLSETFDNLSLK--A	793	
Db	226	REVGRPEPEYQHSRSPHSS-----QSRNCSQRLA	259	
QY	784	SHRSKQHKOSLYGDFVDTNRHDDNRSDNFNTGNMTVLSPLYLNTVLPSSSSSSRGLDS	843	
Db	260	SOAREPTSPSGG-----SRSSSSSDSISS-----SSSNTDSSDS	298	
QY	844	SRSEKSLRERIGLGNVHPATENPGTSKRGQLQISTTAAQIAKWE-----	893	
Db	299	SSTASDSDPAR-----SVGSAAPAPTQSLLSLEKDEPRKFGIK	339	
QY	894	-----VSAIHTSOED-----RSGSGTT--ELHCVTDERNAL-----RRSSAAHTSNT	934	
Db	340	VONLPVRSIDTSLKDLGFHEFKFGKTVSVQIHGASEERYGLVFFRQOEDKALTASKG	399	
QY	935	YNF-----TKSEN-----SNFTCSM-----	949	
Db	400	KLFGMQIETVAVGPETESFNPRDLDERIDFHPKATRTFLIGNLEKTTTHDLRNIF	459	
QY	950	-----PYAKLEY-----KSSNDSLSVSSSDGYGKRGQMKP	981	
Db	460	QRFGVIDIDIKVNGVPOAFQYCDIASVCKAKQNDGEYLGNNRLKLGFK--SMP	516	
QY	982	S-----IESYSEDDSK-----FCSYQY-----PADLAHKTHSAN	1012	
Db	517	TNCVWLDGLSSNVSDQYLTRHFCRYGPVWVDFRLKGMALVLYSEIEDAQAQVETKGR	576	
QY	1013	HMDNDGELD-----TPINYSKYSDEQ-----NSGROSPSONERW---ARP	1052	
Db	577	KIGKNKIKVDFANRESOLAFYHCMEKSGQDMRDFYEMLTERRAGQMAQSKHEDWSADAQS	636	

Db 2571 -----PPCDPKH-----ADTQASE----- 2597
Qy 2720 NRLNSFTQVADQK-GTEIKPG-----QNNPVVSTNESSIVERTPPSSSSSKHSSP 2773
Db 2598 -----VFVAADKEKAVPIAPKITSVIRMPVSIDLENSQKITLAKAPQTUTGLVSAL 2651
Qy 2774 SGTVAARVTPNTNPSPRKSSADSTSRPQIETPVN 2810
Db 2652 TGLVNVSLVPVNAKGPVKGVSATLKLGLVSTPAGPVN 2688
RESULT 8
ID_ZEP1_MOUSE
AC Q03172, STANDARD; PRT; 2688 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 40 (Transcription factor alphaA-CRYBP1) (Alpha A-
DE crystallin-binding protein I) (Alpha A-CRYBP1).
GN HIVEP1 OR ZNF40 OR CRYABP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=D8A/2J;
RX MEDLINE=95138112; PubMed=7836383;
RA Nakamura T., Donovan D.M., Hamada K., Sax C.M., Norman B.,
RA Flanagan J.R., Ozato K., Westphal H., Piatigorsky J.;
RT "Regulation of the mouse alpha A-crystallin gene: isolation of a cDNA
RT encoding a protein that binds to a cis sequence motif shared with the
RT major histocompatibility complex class I gene and other genes.";
RL Mol. Cell. Biol. 10:3700-3708(1990).
RN (2)
RP SEQUENCE OF 2024-2688 FROM N.A.
RC TISSUE=Lens epithelium;
RX MEDLINE=90287161; PubMed=1694016;
RA Nakamura T., Donovan D.M., Hamada K., Sax C.M., Norman B.,
RA Flanagan J.R., Ozato K., Westphal H., Piatigorsky J.;
RT "Regulation of the mouse alpha A-crystallin gene: isolation of a cDNA
RT encoding a protein that binds to a cis sequence motif shared with the
RT major histocompatibility complex class I gene and other genes.";
RL Mol. Cell. Biol. 10:3700-3708(1990).
CC -|- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE
CC PALINDROMIC SEQUENCE 5'-GGGAAATCCC-3' IN THE ALPHA-A CRYSTALLIN
CC PROMOTER.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -|- SIMILARITY: STRONG, TO HIVEP2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L36829; AAA98810.1; -.
CC EMBL; L36825; AAA98810.1; JOINED.
CC EMBL; L36826; AAA98810.1; JOINED.
CC EMBL; L36827; AAA98810.1; JOINED.
CC EMBL; L36828; AAA98810.1; JOINED.
CC EMBL; X68946; CAA48762.1; -.
CC PIR; I49477; I49477.
CC HSSP; P15822; 32NF.
CC TRANSFAC; T00007; -.
CC MGD; MGI:96100; HIVEP1.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 5.
CC

DR SMART; SMO0355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 165 170 POLY-SER.
FT ZN_FING 407 429 C2H2-TYPE.
FT ZN_FING 435 459 C2H2-TYPE.
FT DOMAIN 799 802 POLY-SER.
FT DOMAIN 851 854 POLY-PRO.
FT ZN_FING 953 976 C2HC-TYPE (POTENTIAL).
FT DOMAIN 1482 1486 POLY-SER.
FT ZN_FING 2074 2096 C2H2-TYPE.
FT ZN_FING 2102 2126 C2H2-TYPE.
FT DOMAIN 2179 2182 POLY-ASP.
SQ SEQUENCE 2688 AA; 288341 MW; 5EAD46C3A7008BE6 CRC64;
Query Match 2.8%; Score 404; DB 1; Length 2688;
Best Local Similarity 18.4%; Pred. No. 1.3e-07;
Matches 463; Conservative 328; Mismatches 896; Indels 826; Gaps 106;
Qy 1007 KIHANHMDDNDGELDTPIVNSLYKYSDEQLNSGRQSPQSNRWARPKHII-EDBIKQSEQ 1065
Db 6 QTHPRN-LRDKIEBAQKELNGAEVSKVELEAGVGTSESLKGVKRIKVAENHLKIPK 64
Qy 1066 RQSRN-----QSTTVPVVT--ESTDDKHLKFPHFQQQECVSPYSRGANGSETNRVG 1116
Db 65 SPLNPLQTHKQNTPEPPFVLSASESHVK-----HNCVPAKQGR-----QFTKQNG 113
Qy 1117 SNHGINQVNSQLCOEDDYEDDKPTNYSERYSE-EEQHEBEERPTNYS-----IKYNE 1168
Db 114 ETPGTAESSES-----GDLVSPKKTSPHQRSRLRRWRSEGDPTRLSLDGLDQGDSSSS 169
Qy 1169 EKRHVDPQIDYSLXYATDIPS-----SOKQSFSPKSSSGSQSKT 1208
Db 170 SKARTDNSECSPPCSTTPPSYTTAFDVLKAMEPELSTLSQKSSCAIKTEKLNK 229
Qy 1209 EHMSSSESTTPSSNAKRNQL-----HFSSAQSRSGQFQAATC----- 1249
Db 230 VRSPSKLNSLDAPNATSDPLVVEPCPPCTSYFVHVAQTQSEQVAAQCVSHLYSSQD 289
Qy 1250 ----KVSSINQE-----TIQTYCVEDT---PICFRCSSLSLSAEAEIGCNOT 1292
Db 290 HLVPKLSQQNQQLPGLHGTGSLNLTNLTLESTKLEPIYNTAVTSTVGLTSPSTRTQVTPP 349
Qy 1293 TQEADSAANTLQIAIEKIGITRSAEDPVSE---VPAVSQHP-----RTKSSRLQGS 1340
Db 350 HQQMDSVSPLSVSPAS---STQSPPGPIYSSAHVASVVSQSVQEMCSLLLRDQKFKQK 406
Qy 1341 -----SLSESARHKAVERFSSGAKS-----PKSGAQTPKSPPEHV----- 1377
Db 407 YICBYCNACAKPSVLLKHRSHTGERPPCVTCGPFKTKSNLYKHGKSHAHTIKLGLV 466
Qy 1378 -----QETPLMFSRCTSVSLDSFESRSIASSVQSEPC-----SGMVSGIISP 1420
Db 467 LQPEAGGLFSQECPKALSVDHSDIEDSGDEGLADGRQNNPCVKDLQVPQTKVTSNP 526
Qy 1421 SLDLPSDQGNTP-----PSRSKTPPPPPQTAQTKRVPKNAKTAERES 1465
Db 527 ESLEP-----KLIPSNSDHVRGFSQSDRPSQAPTELKPVVHVPMPLKTDCLQVANP 582
Qy 1466 GPKQAAYNAQVRVQVLPDADTLHFEATSTPDGFCSSLSLSALSLDEPIQK-DVELRI 1524
Db 583 NPELPS-----PQSPRDLHVASI-----LSHSASVSLMEDESCHQGDVQISE 626
Qy 1525 MPVQENDNGNETESQPKSNENQEXAEKTDISEKDL--LDDSDDDDIIELEECI--- 1579
Db 627 GKP--DSHSGTAHAQLQRQATQEDPQEQCKLLSPSLGLSTDGYSFSSASDAQVSP 684
Qy 1580 -----ISAMPTKSRKAKKPAQTASKLPPPVARPSQLPVY 1615
Db 685 TPFARTPTTMDPDPKNGGAPGRISAPASALATGKSSVVTGQMKRPPLATKLEERIS 744

Db 733 SYIQD-----ETIPGYSETEQTSIDEEIHDEPEERPAPEPHSTYDLPFGPEGAGPFAS 787
 QY 1038 -----SGROSPQNERWARPKHLEDEIKQSEQRNOSTTYPVYTESTDDKHL 1087
 Db 788 QPADSAVATSKVXGTETELTYTNIAAPLABEEHVSAITTEC-----DKLS 839
 QY 1088 KFPQHFQOECV-----SPYRSGANGS-----ETNRVGSNHGINQ-----NVQSLS 1129
 Db 840 SFATSAVEDQSASLTAPOTEBTGKSSLLDVTISIPSSRTAQTGLDYVPSAGTISPTS 899
 QY 1130 COEDD-----YEDDKNTYSER-----YSEEEQHBEEREPTNYSIK----- 1165
 Db 900 SLEEDGKFPCCDFSVTGESEKKEGIIIGKLSGERAVEEBEETANVEMSEKLCQYQ 959
 QY 1166 -----YNEEKRHVDQPIDXSLKYATDIPS-----S 1190
 Db 960 TPVFSAPGHALPGEPALGEAEERCLSPDDSVKVASPPSPGPSATHPTPHQSPVEKS 1019
 QY 1191 QKQSFSPKSSGQSKSTEHM-----SSSENTSTP-----SSNAKROQLHPSAQSAGQPOK 1245
 Db 1020 EPQDFQ-EADSWGDTKRTPGVGKEDAAEETVKPPEEGTLEKEEKVPP-----PRSPQAOE 1074
 QY 1246 AATCKVSSINQIOTYCVEDPTICFSRCS-----LSSLSAEDDEIGCNOT 1292
 Db 1075 APVNIDEGLTGCTIQLLPAQDXKAIYFRIWEAGEPTGPILAGAALPGGLRTILPOEFGKQK 1134
 QY 1293 TO-----EADSANTLQIAEIKETIGTRAEDPVSEVPAVSQH-----PRTKSRLL 1337
 Db 1135 DEVLRYPRDSRSPEDAESILSVSPSPDTANOEPTKFGCLTEQYLHKDRWPVSPEDT 1194
 QY 1338 QGSSLSSES-ARKHAEVFPSSGAKSPKSGA-----QTEPKS 1371
 Db 1195 QLSLSSESPSKETSLDVSXSKQSPESIGTLQFGLNLGKEEMGHLMQAEOTSHHTAAMS 1254
 QY 1372 PPE-HYVOETP-----LMFSRCTSVSSLDSPESRSIASVQSEPCSG-----MVGSIISPS 1421
 Db 1255 VPEPHAATASPTDGTTRYSAQTDITD-DLDRKSPASSFSHSTPSGNGKYLPGAITSPTD 1313
 QY 1422 DLPDSFGQTMPSPRSKTPPPPTQATQKRE--VPKNKAPTAE-KRESGPKQAAV---NAA 1475
 Db 1314 EHILTPDSFSKSPESLPQPALEDIAIKWEDKVPGLKORTSEKKEPEPDEVOQQDKOT 1373
 QY 1476 VORVOVLPAADT-----LLHFATESTPD-----GFSCSSLSAL-SLDEPFIOKD 1519
 Db 1374 LSHKEVVEPKDAILYKQBALHVKEAVKQDKALEQKRDLEQDKTALQKDALEPKD 1433
 QY 1520 VELRIMPPVOENDNGNETSEOPKESNENQEKYEAKTIDSEKDLDDDDDDIE-----IL 1575
 Db 1434 KDL-----EEKDKALEQKDKIPEE--KDKALEQKDTALEQKDALEPKDKLEQKORVL 1485
 QY 1576 BECIISAMPTKSSR-----KAKKPAQTASKLPPPVVARKPSOLPVVYKLLPSQNRLOP 1626
 Db 1486 EQ--KEKIFEEKDKALDQKVRSEVHEKAPEDTVAEMKDRDLEQTKAPEQK-----HQAOE 1538
 QY 1627 QK-HVSFTFGDDMPRYVCVETGPIINFSTATSLDITIEPPNELAAGEVGRGAQSGBE 1685
 Db 1539 QKDQVSEKKDQALEQKYWALG-----QKDEALEQNIQALEENHQTQ 1579
 QY 1686 KCDTIPTEGRSDTEAOGGKTSVTIPELDDNKABEGDILAE-----CINSAMPKG 1735
 Db 1580 EQESLVQE-----DKTKKPKVLEKSEPKVAMEEKLEALLETKALGULEESLVQEGRAE 1635
 QY 1736 KSHKPRVKIMDQVOQASASSAPKNQLDGKK-----KPTSPVKPIPQNTY-----R 1786
 Db 1636 QEEKYWRGDDVVQEWQETS-----PTREEPAGEQKELAPAWEDTSP-----EQDNKYWRGR 1686
 QY 1787 TRVKNWADSK-NNLNARVFP-----SDNKDSKQNLKNNKSKDFNDKLPNNEDRV----- 1834
 Db 1687 EDVLAEQDTYWRLESCERKVPWPHELDGGGARPHYTEERESTFLDEGPDDEQEVFLREHA 1746
 QY 1835 -RGSFAPDPSPHVTPTEGTPYCFSRNDSLSLDFDDDD-----VLSREKAEURKAKENKE 1889
 Db 1747 TRSPWASDFK-----DFQESSQKGLEVERWLAESFVGLPPEE 1784

QY 1890 SEAKVTSHTELTNSQOSANKTQAIARQPINRGQPKILOKQSTFFQSSKIDPDGGAATDE 1949
 Db 1785 EDKLTSPFEIISPPAS-----PPENVGQRPSPAPQESIP-----DP 1823
 QY 1950 KLONFAIENTPVCFSHNSSLSLSDIDQNNKNNKENPEIKETEPDQSQGSPKFSQAAGYAP 2009
 Db 1824 KLMPH-MKNEP-----TTPSWLADIPW--VPKDRPLPAPLSPAPGPPPTP-----AP 1868
 QY 2010 KSFHVEDTPVCFSRNSLSLSLSDSEDDLLQECISSA-----MPKKKKPSRLKGDNEKHS 2064
 Db 1869 ES-----HTAPFWSGCTP-----EYDSVVAOVAEGAELEGPPSLGKYKXVAGEREE- 1919
 QY 2065 PRNMGILGSDTLDLKDIQRPDSEHGLS-----PDSENPDKWAIQEGAN 2109
 Db 1920 -----GRAEADKSHXSKVPKARKSHATTEPEQTEPEOREPTPYPDERSFOVADIYE--- 1972
 QY 2110 SIYSSIHQAAAAACLRQ-----ASSDSDSILSKSGISLSPHPLTPDQEKFTNKGP 2165
 Db 1973 ---QMLTGLGPACPTREPPLGAAGWPPCLSTKEAAGRNN-----TSAEKELSPISP 2023
 QY 2166 RILKPEKSTLETTKKIESESKGIKGGKVYSIITGKVRNSNISEIQOMKQKLOANMPSIS 2225
 Db 2024 ---KSLQSDTPTF-----SYAALAGPTVPXPYEPGSPME 2054
 QY 2226 RGRTHIIPGVRNNSSTSPVSKKGPPLKTPASKSPSEGTATTSPRGAKPSVK----- 2279
 Db 2055 PSTTPPAVP-----PRADILSKGP-----SPPLNGNILSCSPDRRSPSPKESGRSH 2100
 QY 2280 -----SELSPVARQTSQIGGSKAPSRSGSRDSTSPFAQOFLSRP-----IQS 2323
 Db 2101 WDDSTDSSELEKAREQAE-----KEAQSPSPPHPIPMGSPPTLWPTEAHVSP 2148
 QY 2324 PGRNSISPGRNGISPP-----NKLQOLPRTSPSTASTKSSSGSKMSYT-----SPG 2370
 Db 2149 PLXSHLGXARPSLDFPASAFGFSLEXAP-POXPSPAEPRSPAFCSGLAFSGDRLALAPG 2207
 QY 2371 ---RQMSQOULTKQGLSKNASSIPRSES-ASKGLNOMNGNGANKVKELSRMSTKS 2424
 Db 2208 PPRTRVDEYLEVTAKPSLD---SSULPQSPSPGXPLSN-----LPRPASPAL 2255
 QY 2425 SGSEDSRSEPVLRQSTFTFKEAPSPPTLRKLKEESAFE-----SLSPSSR 2470
 Db 2256 SEGSSSEATTPTVI-----SSVAERFSPSLEAAESEGELDPGMBPAAHXLWDLTFLSPAP- 2310
 QY 2471 PASPTRSQATPVLSPLPDMSLSTHSSVQAGWRKLPNLSPTIEY-----NDG 2520
 Db 2311 PASLDLALAPSLFGDMGDGILPCHLECSAATEKPSPOVFSQVSECAANGPTTETSNNPP 2370
 QY 2521 RPAKRHDIAKSHGESPSRSLFINRSWTW---KREHSHKSHSSLPVSTWRTWRTTSSSSLSA 2576
 Db 2371 XPAP-----AKAENEAAAXPAWERGAWPEGAERSRSDPTXLSPEQVPCVXGSGGPPSSA 2426
 QY 2577 SSE-----SSEKAKSEDEKHVNSISGT-----KQSKENQVSAKGTWRLK--KE 2617
 Db 2427 SPEVEAGPOQCTEPRPHRGELSPFLNPLPPSIDDRDLSTBEVRLVGRGRRRVGGPG 2486
 QY 2618 NEFSPTNSTSTQTSVSSGATNGAASKTILYOMAPAVSKTDEVMVRIEDCP----- 2665
 Db 2487 TTGXPXVPTDETPTTSASDSSGSSQS-----DSDVPPETEPCPSITAEALDSDE 2535
 QY 2666 -----INNPRSGRSPGTNTPTPVVDSVSEKANPNIKDSKNOAKQONVNGS 2710
 Db 2536 DGDFLPVDKXGVSGTHHPRPHGHPD---PPL-----FPQDXRPS----- 2571
 QY 2711 VPMTVTGLNRLNLSFIOVDAPDOQGTETIKPGQNNPVVPSVETNESSIVERTPFSSSSSSKH 2770
 Db 2572 -PPAP-----DVCMADPECLSESGRVERLEKXKVKQGRVORRAP-----GKDKP 2615
 QY 2771 SSPSGTVAARVTPFNYPNPSPRKSADSTARPQIPTFPVNNNTKKRDSKTDSTESSG 2827
 Db 2616 VSPXERLKLRL---QKRSPTGKGSXDRVSRXPKRSRSTXSTQVTPABEKGQHSPPMSKG 2669

897	QY	IHTSQEDRSSGTTTELHCVTDERNALRRSAAHTHSINTNFYKSE---NSNVRTCSMPYAKL	954
1421	DB	1YK:SESQOEFEED-WTSEKN-----DETSETSVLKHSLVNEVPVLA9PDLLS	1472
955	QY	EYKRSND---SLNSVSSSDGYGKRG-QMKPSIESYSEDDSEKFSYGYCPADLAHKHS	1010
1473	DB	EYSEMKQDLIKMTAILTTDVSCKAGSIKYKELVKAEEBPGFPF-----	1516
1011	QY	AHMDNDGELDTPINYSUKYDEQLNSG-----RQSPQ-----NERWAPKHIIE	1057
1517	DB	-----EIVERVKEDLEKNEILRSGTCTDRDESSVQSSRSRGLVLEEWW-----IYS	1563
1058	QY	DEIKOSEQRQSNQSTTYPVYTESDDKELKQPHFGQOECVSPVRSRCANGSETNRVGS	1117
1564	DB	DEIEBEARQKAPLEITEYFC-VEVRIDKEIK-----GKVE-----KDSGLVNYLTDLNT	1613
1118	QY	NHGINQNVQSLOQEDDYEDDKPTNYSE-----RYSEEECHBEEERPTNYSIKYNEEKH	1172
1614	DB	CVPLPKEQLQTV-----QDRAGKKECALAVGRSEKEGKI0PDETOSTOKHQHSLG	1666
1173	QY	VDOPIDYSLYKATDIPSSQKQSPFSKSSGQSSKTEHMSSESSENTSTPSSNAKRONQLH	1232
1667	DB	IKKPVARKUL-----EKQKQ-----KEGLQASAEKA-----EUK	1696
1233	QY	PSSAQRSSQ-----POKAATCKVSSINOETIQTQVCEDTPICFRSCSSLSSLAEDBI	1287
1697	DB	KGSSESLGEDPCLAPEPLTVKATS-----PLIEETPL-----GSIKDKV	1737
1288	QY	GCNOITQEADSANTLOIAIEIKKIGTRSAEDPVSEVPVAVSQH-----PRTKSSRLQGS	1340
1738	DB	XALQKRVEDE-----QKGRSKLPIRVKGBD-----VPKXTHRHPAASPUSLKSEHAPG	1788
1341	QY	SLSSSARHKAHEFFSGAK-----SP-SKSGAQTPKPSPHYVQETPLMFSRCTS---	1389
1789	DB	SPSPKTERHSTL--SSAKTERHPVPVSPSKTEKHSVPSPSAKTERHSPASSSKTEKHS	1846
1390	QY	VSSLDSDFSRSIASSVQSE-----PCSGMVSGIISPSDLPDPSQOTWPPGRS-KTPPPPPQ	1444
1847	DB	PVSPSTKTERHSPVSVSTKTERHPP-----VSPSGKTDKRPVPVSPSGRTEKHPPVSP-	1897
1445	QY	TQTKREVPKNKAPTAEKESGEKQAAVNAOVQVLPDADTLLHFATETSPDGFSCSS	1504
1898	DB	-GRTEKRLPVSPSGRDTKHQ-----PVSTAGKTEKHLPVSPS---GKTEKQP-----	1940
1505	QY	SLSALSLDPPIQOKVELR-IMPVQO--ENDNGNETESEQPKESNENQBEAKET-IDSE	1560
1941	DB	PVSPSTKTER-IEETMSVRELKMAFGQGDPSKHTGLFEHKSQKQFOEKGKVRVEKE	1999
1561	QY	K-DLLDDSDDDIEILIEECIIISAMPTKSRKAKKPAQTAS-----KL	1601
2000	DB	KGPILTQREAAQTENQTIKRGQBLPVTGTAESKRGVRVSSI0VKKEDAGGKEKVLHSHKI	2059
1602	QY	PPPVARKP-----SOLPVYKLLPSQN-----RLQPKQKHS-----FTFGDDMPR	1640
2060	DB	PEPVQVSPBEESHRESEVPKEKWADQGDMDLQISPDRTKSTIDFSEVIKQELDNDKYQO	2119
1641	QY	VYCVETGTPINFSTATSLDUTIESPPN-----ELIAGGEVGRGQASGEFEKRDII	1690
2120	DB	FRLSEET-----EKAUHLQOVLTSPPNTTFFLDYMKDEFPLALSGLGALDGSSE---	2172
1691	QY	PIEGRGTDEAQGGKTSVT--IPELDNKAEGDILAEACINSAMPKGKHKPRVKKIMD	1748
2173	DB	KNEG-----VAGSPCCSLMEGTFOIISSESYKHGELAETPETS-PESLAFSP---KXSEE	2223
1749	QY	QVOQASASSAPNNQIDGKKKXPTSPVKPIPONTEYR--TRVRKNADSKNNLNARVPS-	1806
2224	DB	QTGETKESTKTETTTTIRSEKEHPTT--K0ITGGSBERGATVTDESETSPESFQKEATLQ	2281
1807	QY	DNKDSKKQNLKNNSKDFNDKLPNNEORVRGS-----FAFDSPHHYTPIEGTPYCFSRNDSL	1862
2282	DB	SPKDTS-----PKQDDCTGCSVALAKETP-----TGLTEBAACDEGQRTF	2323
1863	QY	SSLDLF---D-DDDDYDLRKAELRAKENKESEAKVTSHTELTNSQOSAKNTQIAKQPIN	1919

2324	QSSAHKTQDSEAQESTATSDTKALPIPEASVKTTDTGTE-----SKPQVIRSPQG	2376
1920	RQQXPILQKOSTFPQSSKDIPDRGAATDEKLOFAIENTPVCFSHNSLSLSDIDQEN	1979
2376	LELALP-----SRDSEVLGAVADDSL-----AVSHKDSLEASPVLEDNS	2414
1980	NNK-----ENEPKIKETEPPDS--QGEPSPKQ--ASGYAPKSFHYEDTPVCFSRNSSLSSLSI	2032
2415	SHKTTPDSLEPLSPKSCRDSLESAPVEPKKAGIFPSSHFDI--PAAVAKT-----	2463
2033	DSEDDLLOECISSAMPKKKKSRLLKGDNEKHSPRMGGILGEDLTLDLKDIORPDSEHGL	2092
2464	---ELLTEVASVR-----SRLLRDPDGSA-----EDDSLEQTSLMESSGKSPL	2503
2093	SPD-----SINFWDKAIQEGANSITVSLHQAAAAACLSRQASSSDSDSILSLKSGISLGSFPH	2149
2504	SDPTESSEVSVEYTPKTTD--VSTPKPAVIECAEEDDSENG-----	2544
2150	LTPDOEEKPFTSNKGRILPKPEKSTLETN-----KIESESGIKGKKVKYLSLITGKVR	2204
2545	---EKKRFTPEE-----EMFKWYTKIKMFDELEQEAQ---QKRDYK-----	2579
2205	SNSEISGQMKQPLQANMPISRGRTWIHIPGVNRNSSSTSPSVKGGPPLKTP---ASKSP	2261
2580	---KEPKQ-----EESSSDDPDADCSVDDEPKHTGSGED	2612
2262	SEGQATTPSRCAKSPKSELSPVARQTSQIGSSKAPSRSGSRDSTPSRPAQCPLSRPI	2321
2613	ESGVPVILVTSERKVVSSSESPELAQLKKGADGLLP-----EPVIR--	2655
2322	QSGRGNSISPGRGISPPNKL--SQLPRTSPSTASTKSSGSGKMSY-----TSPGRQM	2373
2656	-----YQPPSPFLSSMDSNSPEEVQFPVVSQYTFKMNEDTOEBPGK--	2699
2374	SOONLTKQTLGSLKNASSIPRSBSAKGLNOMN-----NGNGANKKVELSRMSTK-	2423
2700	SEEEKDSSEHLAEADRAHV--STBAERDSDYDKLARDTDQPKICDHGCE---AMEPSSARP	2755
2424	-SSGES-----DRSRPVLVROSTTIKEAPSTLRRKLEESASFELSPPSRSPASTRSQ	2478
2756	VSSGLQSPGDDVDQPVIIYKESL-----ALQGTHEKDTGEELD--VSRAESP--Q	2803
2479	AQTPVLSPLPMSLSTHSSVQAGWRKLPPLNLSPTIEYNDGRPAKRHDIARSHSSPSR	2538
2804	ADCP--SEFSFSSSLPHCLVSEG--KELEDEISATSSIQKTEVTKTDE---TFENLPKD	2856
2539	LPINRSGTWKRHRSHXSSLLP-----RVSTWRBTGSSSSILSASSSESXAKSE-DE	2589
2857	CP--SQDSSITQDTRFSDMDVPVSDILAENDEIYDPOITSPYENVPSQFFSSEBSKQTDA	2915
2590	KHVNIGI-----SGTKQSKENOVSAKG-----TWKRKIK	2616
2916	NHTTGFHSSEVSVTITSPVEDVWVASSSSGTVLSESNFEGODIKVESQLESTLWEMQS	2975
2617	E---NEFSPNTSQTQVSSGATNGAESKTLIYQMAPAVSKTE---DVW-----	2658
2976	DSVSSSEFTMGATTIV---GEOISKVI-----ITKIDVDSWSSEIREDDAEFA	3024
2659	-VRIEDCPINNRSGRSPGTGNTPPVID--SVSEKANPNIKDS-----KDNQAKQNVGNSV	2711
3025	RVKEEQKIFGLWDRQSGTTPDTPPARTPTBEGTPTSEGNPFLFOEGLFEMTRSGAI	3084
2712	PMRTVGLNRLNSFTQVDPADQKGT---EIKQON--NPVPVSETNES-----	2754
3085	DMTKSYADESFHFQIQGESREETLSEVDKEGATGADPLPSETSLESLSKSKTVD	3144
2755	-----SIVERTPSSSSSXHSSPSGTVAARVTPFNYNPSPRKGSADSTSRAPQIPT	2807
3145	EADLLPDSVSEEBEIPASDAQLNSQMGISASTETP-----TKEAVSVGTXLPT	3194
2808	-----PVNNNTKGRDSTKDSTE	2824

```

3195 VQTGDIPLSG--VKQISCPDSSE 3216
DR PFam; PF00084; sushi; 1.
DR PFam; PF00193; Xlink; 2.
DR PRINTS; PRO1265; LINKMODULE.
DR PRODOM; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 2.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
DR Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 26 POTENTIAL
FT CHAIN 27 3562 VERSICAN CORE PROTEIN.
FT DOMAIN 27 143 IG-LIKE V-TYPE.
FT DOMAIN 166 243 LINK 1.
FT DOMAIN 264 345 LINK 2.
FT DOMAIN 3254 3290 EGF-LIKE 1.
FT DOMAIN 3292 3328 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3341 3455 C-TYPE LECTIN.
FT DOMAIN 3460 3518 SUSHI.
FT DISULFID 44 129 BY SIMILARITY.
FT DISULFID 171 242 BY SIMILARITY.
FT DISULFID 195 216 BY SIMILARITY.
FT DISULFID 269 344 BY SIMILARITY.
FT DISULFID 293 314 BY SIMILARITY.
FT DISULFID 3269 3269 BY SIMILARITY.
FT DISULFID 3269 3278 BY SIMILARITY.
FT DISULFID 3280 3289 BY SIMILARITY.
FT DISULFID 3296 3307 BY SIMILARITY.
FT DISULFID 3301 3316 BY SIMILARITY.
FT DISULFID 3318 3327 BY SIMILARITY.
FT DISULFID 3334 3345 BY SIMILARITY.
FT DISULFID 3362 3454 BY SIMILARITY.
FT DISULFID 3430 3446 BY SIMILARITY.
FT DISULFID 3461 3504 BY SIMILARITY.
FT DISULFID 3490 3517 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 948 948 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 485 Missing (in isoform V1).
FT 3562 AA; 388078 MW; 9BC565E88C1602D2 CRC64;
SQ SEQUENCE

```


Db 2164 QSGBEATLTSTKASVPTGKPEQOQVGRKTVSMPSAVHAYTABNELVSTSEHDI 2223
QY 1941 PDGAATDEKLQNPAL-ENTPVCFSNSLSUSDIDQNNKNEPIKETEP--DSQG 1997
Db 2224 TSLQTVTDTMEBEKAANELVTTSFATNPLS-----EDVHSEWDRP-REILFKAIESG 2276
QY 1998 EPK-PQASGYAPKSFHVE-----DTPVCFGRNS-----2026
Db 2277 EATEDPFIQTQANHEHVELSVPTIHPHSEENKVEAESDEKILLFPNDRVTEGSAVIER 2336
QY 2027 --LSSLSIDS--BDDLQECISS-----AMPKKKKPRLKGDNKHGS-PRNMGIIGE 2074
Db 2337 KYLSSPTDTEEBELVQNIPTEDIPRLFTLTKPEKPT-----NNELISDPLFSGQSGD 2392
QY 2075 DLTL--DLKDIQRPSDHLSP-----DSENFDMKA-IOEGANIS 2111
Db 2393 EFTVPSVESLAVKETNTLSPFPHPASVGPKLSTDKTVQFESGTDNAGINEIITT 2452
QY 2112 VSLHQAAAAACLSRQASSDSILSKGISLSPFHLTPDOEEKFTSNKGR-----2166
Db 2453 AAELETATYSMATSSPALBESSHNSKDKDITHYFLVIEDPYNKEMDHRGEGTSR 2512
QY 2167 ILKPEKSTLETCK-----IESKIKGKKVYKSLITGKVRSEISGQ--2212
Db 2513 LPTFGDVLSEESHMLTDDVTVPVSLSTPYLEMCKSLATSKMPRVLFPESGEGS 2572
QY 2213 -----MKQLQANPISIRGRTMIHIPGVNRSSSSTSPVSKG 2250
Db 2573 GWDGVDSFAPDILTHTAFSVMVELTASS-----HIPGVSEVMIT-----2615
QY 2251 PPKTPASKSPSEG-QTATTPRGAKPSVSELSVPARQTS-----QIGSSKAPSRG-SR 2305
Db 2616 -----HVPFGSQTVIT---GLASLFTPEEKEIVANRTAADPKGTSEELTSDTGM 2664
QY 2306 DSTPSRPAOPLRPIQSPGRNIGISPNKLSOLPRTSSPSTASTKSSGSKMS 2365
Db 2665 DIIPVDDRHLVINSVYGDILBERLOI-PSEKTIIDMDHKSMPEDIIISVQTMEN 2723
QY 2366 YTFGRMQOONTKQ-----TGLSKNVAISIPSEASKG-----2400
Db 2724 LVIRSTQVDDNKAEDKYDILNSTVEENSGDNLSTLTSIQPSSEVYTAGHGP 2783
QY 2401 LNMNNGANKKVELSRMSTKSSGSESDRSPVLVROSTIKEAP---SPTLRKLE 2457
Db 2784 LVKDLGGVAMOFATETLIT-----VLNGLGIFLPTVPLVSPHMEHESK 2830
QY 2458 ESASFES-----LSPSSRPASPTRSOQTPVLSPSLDMSLTHSV--QAGGWRKLPP 2509
Db 2831 ES-EFEAKHIGRTSTDDVVEPVTS-ANNQVITDQKTMISIGFSGMQOESGDKKPMIP 2888
QY 2510 NLSP--TIEYNDGPAKHDIASHSPSLRNLNSGTLWKRSHKSSSLPRVSTWRET 2567
Db 2889 SLTPDLTMEKALTDTDFVSMVTTQSMSCATVSSSEBKHS-----TYVMT 2939
QY 2568 GSSSILSASSESEKAKSDEKH-----VNSISGTQSKENQVSAGK 2610
Db 2940 KSASTEYEETDSVLSNVSQNKSSVTVLVNGVSKYVEVIIPTSSAKDSQSDHSSDG 2999
QY 2611 TWKIKENERS-----PNSTSTQTVSS--GATNGAESTLIYQMAPAVSK-----TDDVV 2659
Db 3000 TFEVSDMAATYKPPPTDITVLSLVFSPSESESISTESTPHFNKVTIERSSETES 3059
QY 2660 RIEDCPT--NNPRSGRSP-----TGNTPPVIDSV-----SEKANPNIKDS 2697
Db 3060 SVNDLIIENATVSGDSFSDHYDPTAFWNFGERTSTDPKLSLIEVFESRVRKNSQES 3119
QY 2698 KDNQAKQVNGVSPMRTVGLNRLNSF-IQVDAPOKXGTEIKPGQNNVPVSETNESSI 2756
Db 3120 DRSTERERPLSGAP-----VSDSPNIEVGFKPQOEAVTMTLSTSLDLRSLETQSAL 3174
QY 2757 V-----ERTPFSSSSSKSHSPSGTVAARVTPFNVPNPSRKSADSTS-----AR 2801

Db 3175 LGPLLQOQEITTISSNIATNTAPGN-----NPYSNEQSTISSSELLNTIELVTS 3223
QY 2802 PSQIPTPVNNNTKKRDKSTDTSESSGTQSP 2831
Db 3224 SPSLPEVTNGSDFLIGTSVGSVEGTAVQIP 3253
RESULT 12
PCLO RAT
ID PCLO RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6: Q9JTLI;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.",
RL Neuron 25:203-214 (2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4689; ASP-4674;
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP ALA-4694.
RX MEDLINE=21181819; PubMed=11285225;
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.",
RL EMBO J. 20:1605-1619 (2001).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AF138789; AAF07822.2; -
CC EMBL; AF227534; AAF63196.1; -
CC HSSP; P04410; 1A25.
CC GO; GO:0045202; C:synaptic junction; IDA.
CC GO; GO:0005509; F:calcium ion binding; IDA.

DR GO: GO:0005544; F:calcium-dependent phospholipid binding; IDA.
 DR GO: GO:0005522; F:profilin binding; ISS.
 DR GO: GO:0016010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO: GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro: IPR000008; C2.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR008899; ZnF_piccolo.
 DR Pfam: PF00168; C2; 2.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF05715; ZF_piccolo; 2.
 DR SMART: SM00239; C2; 2.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS00499; C2 DOMAIN 1; 1.
 DR PROSITE: PS00004; C2 DOMAIN 2; 2.
 DR PROSITE: PS0106; PDZ; 1.
 DR PROSITE: PS0106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491
 FT ZN FING 523 547
 FT ZN FING 1010 1033
 FT DOMAIN 2351 2362
 FT DOMAIN 4442 4536
 FT DOMAIN 4653 4752
 FT DOMAIN 4968 5059
 FT VARSPLIC 4876 4880
 FT VARSPLIC 4881 5085
 FT MUTAGEN 4668 4668
 FT MUTAGEN 4674 4674
 FT MUTAGEN 4688 4688
 FT MUTAGEN 4688 4688
 FT MUTAGEN 4689 4689
 FT MUTAGEN 4689 4689
 FT MUTAGEN 4690 4691
 FT MUTAGEN 4692 4693
 FT MUTAGEN 4694 4694
 FT SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
 Query Match 2.7%; Score 390; DB 1; Length 5085;
 Best Local Similarity 20.0%; Pred. No. 8.9e-07;
 Matches 386; Conservative 241; Mismatches 775; Indels 532; Gaps 81;
 QY 1187 IPSGKQSFSSKSSGSSGSSKTEHSSSSSENTSTPSSNAKRNOLHPS--SAQSRSGQPQ 1244
 DB 67 VPPAAEFSMHRKQELDSSQAPOPGKPPDPGRPTQGLSKSRITDIFRSEQLGRSP 126
 QY 1245 KAATKVSINQETITQYCVBDTPICFRCSLSLSAB-----DEIGNQNTQPADS 1298
 DB 127 STISLKESKSRDTDFKEEYKSNMPPGFFDVNPLSAVSVNKNPNFOLLSDSEASQ-E 185
 QY 1299 ANTLQIAIEIKIGTRAEADPV---SEVPANVSOHPRTKSSRLQGS---SLSESARHKA 1352
 DB 186 TKQKVVOKEQKSEGMKAPLOQSPAPIPKQGVKEVIQDSSSPKSVSSQAEKVPK 245
 QY 1353 EFSGAKSPKSGAQTPKSPPEHYVQETPLMFRSCTSVSSLDSPESRSIASSVQSEPCSG 1412
 DB 246 QAPGTGKPSQSPAQTEA-----QASGPKPVAQP--G 277
 QY 1413 MVSGIISFDLPDSQGTMPBSRSKTPPPPTQTAQTKREVPAKVA---PTAKRESGPKQ 1469
 DB 278 SAKATVQPGPAKSPAQ--PAGTGKSPAQP-----AKTGGQAQGLEKTSSSQQPGFKS 329
 QY 1470 AAVNAAVQVRVQLP-----DADTLLHFAETSTPDGFCSSSLSALSLDEPFIQKVLE 1523

DB 330 LAOTPGHGKPLGFKVSPAQPGTAKHPAQPGPQ-----TAAKVPGF----- 372
 QY 1524 IMPVQENDMNGNETESQPKESN-----ENQEKAEKXITIDSEKLLDDSDDDDD 1572
 DB 373 TKTPAQSGGFG-KHPAQPGPTKPSQOPIPAKPPQCPQVATKTQPOQ----- 419
 QY 1573 EILLECIIISAMPTKSSRKAKKPAQTAKLP-----PFVARKPSQLPVYKLLPSQNL 1624
 DB 420 -----SAPAKPQPPQAPAKPQPPQPTPAKPPQPPPTPAKPPQ-----PTATKP 465
 QY 1625 QPKHVSFTPGDDMPRVYCVETPIINFSTATSISDLTIESPNELAAGEVGRGAQS--- 1681
 DB 466 QPOPTATKPHHOOPGLAKPSAQOPTKSIQTVGRPLQPPPTS-AAQTPAQSLSTICP 524
 QY 1682 --GEFERDRTIP-----TEGRSTDEAQGGKTSSTVITPELDD-----NKASEGDI 1723
 DB 525 LCNTTELLHIPEKANFNTCTEQSVCSLCGFNPNPHLTIKEWLCNLCMQQALGGDL 584
 QY 1724 LAECINSAM--PKGSHKPFVRVKINDVQVQAASASSAPNKNQ-----LDGKK 1769
 DB 585 AAAPSPSQPTPKAATAFTATASKSPVPSQASPKPEPSKQSPKALESKKPEPKPP 644
 QY 1770 --KKPTSPVKP-----SSLDPDDDDVLSREKALRAKEN 1778
 DB 645 EPKKPEPKPPLVKQPTLHGPTATAPQLPVAEALPEPAPPEPSGPIPEQAKAPVGD 704
 QY 1779 -----IPONTYRTRVRKNADSKNNLNAERVSNDKSKKNLKNKDFNDKLPNEDRV 1834
 DB 705 VEPKQPMETRADIQSSSTTKPDILSSQVQAQVKTASPLKTDSAKPSQSPPTGEKT 764
 QY 1835 RGSFAPDSPHYTPIEGTPYCFERNDSL-----SSLDPDDDDVLSREKALRAKEN 1887
 DB 765 -----TFLDSKAMPSPASDSKIIISQPGPSESKDPKHIDPIQKDEPKKA-QP 811
 QY 1888 KESEAKVTSHTELTSNQQSANKTQAIKOPINRQPKILQKS-----TFPQS 1936
 DB 812 KGSPKPEKPVKSGSPPTSGTRPTAGAAPPSQPPKP--QEQSRFSLNGLGITDAPKS 869
 QY 1937 SKDIPRGAATDEKLNQFALENTPVCFSHNSLSLSLSDIDQENNNKENEPIKETE--PPD 1994
 DB 870 QPTTPOE--TVTKLFGFGAS-----IFSQASNL--ISTAGQQGPHPTQPAEFSKQAPT 921
 QY 1995 SQGEPSKPOASGYAPKSFHVEDTPVCFS--RNSLSLSLSDSDLLQECISSAMPKKK 2052
 DB 922 SQ-----SPAAQGPASTGQLPPAPAKATAVKEAKAAEAENLESKEPQAPTAKTEKDK 977
 QY 2053 PSRLK-----GDNEKHSRNMGGIIGEDLTLDLK--DIQRPDSHGLSPDSSENFDKA 2103
 DB 978 PPPAKVGGKPPSPBPBAVPAHK-----PDKTTKPKPACPLCRTELNLG--SQEPNEN--T 1029
 QY 2104 IQEGANSI-----VSSLHQAAAAACLSROASS-----DSDSILSKSGISLGSPP 2148
 DB 1030 CTECKNOVCMNGFNFTPHLTIQEWLCNLCQRAISQGLGDMGNKMPAPSP--PKASPM 1088
 QY 2149 -----HLTP-----DOEKPFTSNKGRILKPGKEKSTLETTKIES 2185
 DB 1089 PAPAEPSSOKTPTGTQVKGKKEAGTEAEKVPPEKETASIEKTPPWVTTDQKLESE- 1147
 QY 2186 KGIKGGKVKYKSLITCKVRSNE--ISGQVKQ-----PLQANPISIRGRTMIHP 2234
 DB 1148 -----GKKKGVSALEPKKPEEKEAISADKKERKPPAEKPPLEKPP----- 1190
 QY 2235 GVRNSSSTSPVSKGPPLKTPASKSPSEGTATTSPRGAKPSVKSELSVPAQTQSIGG 2294
 DB 1191 -----IPVDKKLPPEAKPLS---SEGE-----KHEILKAHVQIPEEP 1226
 QY 2295 SVKAPSRGSRDSTPRPAQQLSRPIQSPGRNNSISPGRNGLSPNNKLSQLPRTSPSTA 2354
 DB 1227 TGVAAKAGEE-----QPDSP-----EALPGATPLTPKAGEKERAQAQAE 1272
 QY 2355 STKSSGSKMSY-----TSPGOWSQNLTKOTGLSKNA-----SSIPRS 2394

DR GO: GO:0005509; F:calcium ion binding; ISS.
 DR GO: GO:0005544; F:calcium-dependent phospholipid binding; ISS.
 DR GO: GO:0005522; F:profilin binding; ISS.
 DR GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO: GO:0016080; P:synaptic vesicle targeting; ISS.
 DR InterPro: IPR000008; C2.
 DR InterPro: IPR001565; Synaptotagmin.
 DR PRINTS: PR00360; C2DOMAIN.
 DR PRINTS: PR00399; SYNAPTOTAGN.
 DR SMART: SM00239; C2; 2.
 DR PROSITE: PS00499; C2 DOMAIN 1; 1.
 DR PROSITE: PS00004; C2 DOMAIN 2; 2.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 Repeat; Alternative splicing.
 FT NON_TER 1
 FT DOMAIN 400 465
 FT ZN_FING 499 523
 FT ZN_FING 969 992
 FT NON_CONS 1010 1011
 FT DOMAIN 2300 2325
 FT DOMAIN 4391 4442
 FT DOMAIN 4544 4633
 FT DOMAIN 5031 5121
 FT VARSPLIC 4404 4404
 FT FT
 FT VARSPLIC 4534 4534
 FT VARSPLIC 4576 4576
 FT VARSPLIC 4757 4761
 FT VARSPLIC 4762 5147
 FT SEQUENCE 5147 AA; 563537 MW; CD5D84390498DC3 CRC64;
 Query Match
 Best Local Similarity 19.4%; Score 389.5; DB 1; Length 5147;
 Matches 508; Conservative 313; Mismatches 959; Indels 837; Gaps 122;

718 IAMGSAALNLMANRAKYKIDANIMSPGSSI-----PSLHVRKQKALEALD 765
 1 IPAGNEADLSQI--SEERRQIAVMSRAQGLPKGSVPAAAEPSMH-RKQ-----ELD 52
 766 AQHLSSETFDNIDNLPKASHRSKQHKOSLYGVDFDTRHDDNRSDNFTGNMTVLSPY 825
 53 SSH-----PPKQSGRPP-----DGRPAQPLGSKRRTDTF 83
 826 LNTTVLPSSSRGLSDSRSEKORSLEIRGILGNVHPATENPG-TSSKRGLQISTTA 884
 84 RSEQLPGRSFSTLSKESKRTD-----LKEHKSSMMFGLSEYNALSAVSV 133
 885 AQIAKWEVSAIHTSQEDRSSGSTTELHCVTDE----- 918
 134 VNKENPFLDISSEASQEE-----TTKQKVQKEQKPGEGIIKPELQOQPPKPIPKQOQ 188
 919 --RNALRSSAAHTSNYNTKSENSNRKTCMP-----YAKLEYKRS 959
 189 PGRDPLQDGTGPKSIS-----SQQPEKISQPPGTGKPIQGTPTQDTHAKLPLQRD 241
 960 SNDSLNSVSSDGYKRGQ-MKPSIESYSEDESKFCSYGOVPADLAHKIHSANHMDND 1018
 242 A--SRPQTQADIV--RSESVKPSLPSKSPPI-----OQPTGKPPAQOQHEKSKQP 290
 1019 GELDTPIPNYS-----LKYSDQLNSGRQSPQNERWAPKHIIEDEIKSQSORSNQST 1073
 291 GPAPKPPAQSGLTKPLAQOQGTVPVPPVPPGTGTPKPAQP---LGPAPKPPAQOQSEKPS 347
 1074 TYPVYTESTDHLKFKQHFQOQECVSPYRSGANGSETNVAAGNINQVNSGLQED 1133
 348 EQP-----GPKALAQPPGVGK-----TPAQOFGAKPPTQOVGT-----PKPLAQQP 389

QY 1134 DYEDDK-----PTNYSEYSEBQHEBERFTNYSIKNEEKRVHDQIDYSLKYATDIPS 1189
 Db 390 GLOSPAKAPGTTTPTAQTKPSPQOPGSKTPEPPQ--QFGPAKPSPOQ-----GSTKPPS 441
 QY 1190 SQKQSFSSXSSGQSKTEHSMSSSENTSTPSNAKRONQLHPSSAQSSGQPOKAATC 1249
 Db 442 QQPGS---AKPSAQOQSPAK---PSAQOFTKPVSTQFGKPLQPPVTPVSPAKAPP----- 490
 QY 1250 KVSSINQETIQTVCYVEDTPIC-----FSRCSLSLSLSAEDEICGNQTOE 1295
 Db 491 -----SQLEPKTIC-----PLCNTTELLHVPEKANFNTECTQTTVCSLCGFNPPLHTE 541
 QY 1296 ADSANTLQIAIEIKELIGTRSAEDPVSEVPAVSHBRTKSSRLOGSLSSSARHKAPEFS 1355
 Db 542 AKWLCNLN-COMKRALGGDLAPVPSPOPKLKTAPVTTTSAVSKSPQOQTSPK---K 596
 QY 1356 SGAKSPSKGAQTPKGPPEHYVQETPLMFSCRCTSVSLDSFESRSTASSVQSPSCGMVS 1415
 Db 597 DAAPKQDLSKAPEKPKPPLVKQPT----- 623
 QY 1416 GILSPDLPSQGTMPPSSEKTPPPPPQTAQTKREVKNKAKTAEKRESGPKQAANAA 1475
 Db 624 G--SPSAKAKQP-----PEADLSKPAPEKPSVPSE--QDKAPVADKKQPKM----- 669
 QY 1476 VQRVQVLPDADTLLHFATESTPDGFCSSSLSALSILDEPFIQKDVLRIMPVPPQENDNGN 1535
 Db 670 -----VKPITLVSSTATTXPDIPSSKVQSOABEKTTPPL----- 705
 QY 1536 ETSEQPKES-----NENOEKAEKTIIDSEKLLDD 1566
 Db 706 KTDSAKPSQSPPTTGEKVTFFDSKAIPRASDSKIIISHPGPSSSESGQKV----- 756
 QY 1567 SDDDDIEILIECIISAMPTKSSKAK-----KPAQTASKLPP--PVARK---PSQLPV 1614
 Db 757 ---DPVQKKEE-----PKAQTKMSPKPDAPKMPKSGTPPPGPRTAGVTTPQOSPK 807
 QY 1615 -----YKL-----LPSQNELQPKHVS-----FTPGDDMPRVY 1642
 Db 808 PQBSRRFSLNLGSIITDAPKSQPTTQETVTGKLFQGFASIFQASNLISTACQPGPHSQ 867
 QY 1643 CVRGTPINFTATSLSDLTIESPPNELAAGEVYRGGAQSGEFKRTDITPEGRSTDEAQ 1702
 Db 868 SGFGAPWKQAPAS-QPPTSQGGPK--STGQAPPAKSIPIVKETKAPAAEKLEPRAEQ 924
 QY 1703 GKTSSVT-----IPELDDNKAEGDILAEICINSAMPK--GKSHKPPRVKIMDQVQASA 1755
 Db 925 APTVKRTETETKPPKIDSKS---LTAEPQKAVLTKLEKSPKPESTCPLC-KTELNI 979
 QY 1756 SSSAPNKN-----QLDGGKKKPTSP-----VKP 1778
 Db 980 SKDPFNFNTCTECKNVQCNLCGFNPTPHLTENCQTORAISGQLGDIRKMPAPSGPKASP 1039
 QY 1779 IPQNTYVTR-----VRKNADS-----KNLNAER-----VFSNKNDS 1811
 Db 1040 MPVPTSSSQKTAVPQVKLVKQEQEVTEAEKVLKVKETLSMEKIPMWTTTQKOE 1099
 QY 1812 KQNLKNNKDFNDKLP-----NNEDVRGS-----FADSPHHYTPIEG--TPYCF 1856
 Db 1100 ESKLEKDKASALQEKPLPEEKLIPEEEKIRSEKKPLLEKKP---TPEDKLLPEAK 1156
 QY 1857 SRNDSLSDDFDDDDVDLSREKAELEKA---KENKESAK---VTSHTELTSNQQSANK 1309
 Db 1157 TSAPEQKQDHLKSQVIAEEKLEGVAVKTVQEGKQPTQKMEGLSPGTFQSLFKEDDKT 1216
 QY 1910 TQIAKQPINRGQPKFILOKQ-----STFPQSKDIPDRGAA----- 1946
 Db 1217 TKTIKEQPQPCPTAKPDQEKEDDKSDTSSSQPKSPQGLSDTGYSDDGSSSLGILPISLI 1276
 QY 1947 TDE-----KLONFAIENTVCFPSHNSLSL-----LSDIDQENNKENEP--- 1986
 Db 1277 PTDEKIDILKGLKDKDSQSSSSPSSDLAKLSTVLISLEAQASTLADEKSEKKTOPHEV 1336
 QY 1987 -----IKETFPDPSQGEPSK-----PQASGVAPKSFHVE 2015

Db 1337 SPEQPKQDKQKTSQSLSETLEITISEEBIKESQBERKDTFKKDSQODIPSSKDKHEKSEFVD 1396
Qy 2016 DTPVCFSRNLSLSLSDSDLLQECISAMPKKKPSRL-KGDNKEHSPRNMGGILGE 2074
Db 1397 DIT---TRRPYDSVSESE---NSPVQKARTSVGSSSDSEYKQEDSQSGSEE 1447
Qy 2075 DLTLDKDIQRPDSHGLSPDSFNPKAIQE-----GANSIVSSLIHQAAA 2120
Db 1448 EDPIRQIEMSADEADASGEDDEFIRNOLKETSSTESQKBEETKGGKITAGKHR-- 1505
Qy 2121 AACLRQASDSDSILSLKSGISLGFPHLTPDOEKPTSNKPRILKPGKSTLETCK 2180
Db 1506 ---LTRKSTSIDEDAGR---HSHDEDEAF--DESP-----ELKRETKS 1545
Qy 2181 IESKSGIK--GGKVKYKSLITGVRSNBEI-----SQMKQPL-----QANPMSIS 2225
Db 1546 QSEELVVTGGGLRRPKT-----IELNSTIADKYSAESQKTSLYFDEEPELEMESLT 1600
Qy 2226 ---RGTWIIH---IPGVNRS-----SSTSPVKKGPPKTPASKSPSEG- 2264
Db 1601 DSPEDRSRGEGSSLSHASTPTGTSPTSVSLSDEDSPSHKKGSQKQKARHRPHGP 1660
Qy 2265 --OTATTPRGAKPSVKSELSPVARTQTSIGGSKAPSRSGSRDSTPSPPAQOPLSRPTQ 2322
Db 1661 LLPTIEDSSEEBELREBEELLKQEKQRETEQOQRKSSKSKKDKDELRAQRRRERPKT 1720
Qy 2323 SPGRNSTSPGRNGISPPNKLQ-----LPTG-SPSTASTKSGSGKMSYTSIPGRMSQ 2375
Db 1721 PP--SNLSPIEDA-SPTEELRQAEMELHRSSCSYSPSIEDPEG--FEISPEKIIIV 1775
Qy 2376 QNLTK-QTGLS---KNASSIPRSEGASKL-----NOMNGN-----GANKKVEL 2416
Db 1776 QKVTKLTAVLSLSPDDEQIMKESQKALKSAEMYEEMHKTXYKAFPAANERDEV 1835
Qy 2417 -----SRMSSTKS---SGSSEDRSE-----RPVLVRQSTFTK 2445
Db 1836 FEKEPLYGGLIEDIYIESLVEDTYNGSVGSLTLTQEEENGFMQKGREKIRLSEQIV 1895
Qy 2446 EAPS---PTLRKLEESAFESLSPSRPASPTSOAQPTVLPS--LDMLSLTHSSVQ 2500
Db 1896 EDFMQKITDQKEPEYELSHSVVP-----QEDIVSSSIIFE-----SHEIVD 1939
Qy 2501 AGGW-----RKLPPLNLSPTIEYNDGRPAKHDIARSHSESPSRLPINRSGFTWKREHSK 2553
Db 1940 LGTMVTSTEEERKL---LDADAAYEE--LMKROQMQLTPGSSPTQAFIGEDMT---ESTM 1991
Qy 2554 HSSSLPRVTRWRTGSSSSILSASS-----ESSEKAKSEDEKHNSISGTQSKENQVYSAKG 2610
Db 1992 DFDMPDASL-----TSSVLSGASLTDSTSSATLSIPDVKITHQFSTEBIEDEYVT--D 2043
Qy 2611 TWRIKE-----NEF-----SPTNSTSQTVSSGATNGASK-TLIYQMAP 2649
Db 2044 YTEIQIHAESLILTYSPSSSATSVPPSDPTSLTSSVSVCTTSSSPITLDSIT 2103
Qy 2650 AVSKTEVDVRIEDCP-----INNPGRSGRPTGNTPP----- 2681
Db 2104 VYTEPVDMTKFEDESEBISSTYPPGSIIDYPEISVSLDRTAPPDGRASADHIVLSLD 2163
Qy 2682 ---VIDSVSEKANPNIKDS-----KD--NOAKQNVGNGSV-----PMTVGL 2718
Db 2164 MASSIIESVVPKPGPVADVTSDLLISEKDPVKAKKFTGNGIILEVLEYADKKELEA 2223
Qy 2719 ENRLNSTIQV--DAPDQKGTIEKPGQNNVPVSETNESSIVERTPFSSSSSS---SKHSS- 2772
Db 2224 ERTKSSLSSETVFDHP-----PSSVIALPMKEQ-----LSITYFTSGTFTGQKPPASQ 2270
Qy 2773 -PSTGTAARVTPFNYPNPSPKSSADTSARPSQIPTP 2808
Db 2271 LPSSGSPSVSSLPAKPRBFRFRSSSLD-ISAQPPPPPPPP 2306

MLL3 HUMAN
ID MLL3 HUMAN STANDARD; PRT; 4911 AA.
AC Q8NEZ4; Q8NC02; Q8NDF6; Q9H9P4; Q9NR13; Q9P222; Q9UDR7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-
lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43)
DE Homologous to ALR protein).
GN MLL3 OR HALR OR KIAA1506.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Petal thymus;
RX MEDLINE=21888622; PubMed=11891048;
RA Ruault M., Brun M.-E., Ventura M., Roizes G., De Sario A.,
RT "MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a
RL chromosome region frequently deleted in myeloid leukaemia.";
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21574953; PubMed=11718452;
RA Tan Y.C., Chow V.T.;
RT "Novel human HALR (MLL3) gene encodes a protein homologous to ALR and
RL to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated
RN with leukemia and developmental defects.";
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22737999; PubMed=12853948;
RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
RA Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
RA Fewell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
RA Ozersky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
RA Kozlowski-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
RA Tin-Wollam A.-M., Abbott A., Minx P., Maupin R., Strommatt C.,
RA Latreille P., Miller N., Johnson D., Murray J., Woessner J.F.,
RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,
RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,
RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
RA Clifton S.W., Chissoe S.L., Marra M.A., Raymond C., Haugen E.,
RA Gillett W., Zhou Y., James R., Phelps K., Iadonato S., Rubb K.,
RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,
RA Baertsch R.A., Brent M.R., Keibler E., Plicek P., Bork P., Suyama M.,
RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
RA Waterston R.H., Wilson R.K.;
RT "The DNA sequence of human chromosome 7.";
RL Nature 424:157-164(2003).
RN [4]
RP SEQUENCE OF 556-3865 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RL The complete sequences of 100 new cDNA clones from brain which code
RN for large proteins in vitro.";
RN [5]
RP SEQUENCE OF 3193-3865 AND 4460-4911 FROM N.A.
RC TISSUE=Placenta;
RA Isozai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.;
 "NEDO human cDNA sequencing project.";
 Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RN SEQUENCE OF 3879-4911 FROM N.A.
 RP TISSUE=Testis;
 RC Duesterhoeft A., Lauber J., Mewes H.-W., Weill B., Wiemann S.;
 RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RL
 [7]
 RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX (ISOFORM 2).
 RC MEDLINE=22311436; PubMed=12482968;
 RX Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J., Kwak E., Barlow N.A., Berger S.L., Chow V.T., Roeder R.G., RA Acores D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C., RA Lee J.W.;
 RP "Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins.";
 RT Mol. Cell. Biol. 23:140-149(2003).
 RL
 CC -1- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a coactivator complex of nuclear receptors, involved in transcriptional coactivation. MLL3 may be a catalytic subunit of this complex, which weakly methylates Lys-4 of histone H3. This is a specific tag for epigenetic transcriptional activation. May be involved in leukemogenesis and developmental disorder.
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine = S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
 CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/RBBP5, alpha- and beta-tubulins, the trithorax group proteins MLL2 and MLL3, and ASH2/ASCL2. Interacts with histone H3.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC -1- Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8NEZ4-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8NEZ4-2; Sequence=VSP_008561, VSP_008562;
 CC -1- TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed by brain and liver. Also expressed in placenta, peripheral blood, fetal thymus, heart, lung and kidney. Within brain, expression was highest in hippocampus, caudate nucleus, and substantia nigra. Not detected in skeletal muscle and fetal liver.
 CC -1- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B and H4, and may have a H3 lysine specific methylation activity.
 CC -1- MISCELLANEOUS: Found in a critical region of chromosome 7, which is commonly deleted in malignant myeloid disorders. Partial duplication of the MLL3 gene are found in the juxtapositional region of chromosomes 1, 2, 13 and 21. Juxtapositional reshuffling of the MLL3 gene has generated the BAGE genes.
 CC -1- SIMILARITY: Belongs to the TRX/MLL family.
 CC -1- SIMILARITY: Contains 1 DHHC-type zinc finger.
 CC -1- SIMILARITY: Contains 6 PHD-type zinc fingers.
 CC -1- SIMILARITY: Contains 1 post-SET domain.
 CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
 CC -1- SIMILARITY: Contains 1 SET domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcements> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AY024361; AAK00583.1; -
 DR EMBL; AF264750; AAF74766.2; -
 DR EMBL; AC006617; AAD45822.1; -
 DR EMBL; AC104692; -; NOT ANNOTATED_CDS.
 DR EMBL; AC005631; -; NOT ANNOTATED_CDS.
 DR EMBL; AB040939; BAA96030.2; -

[illegible]

Db 734 TGLMSEMTTIEGCVK--DV5YQG--GSKISSLSETESSADISKADVSSPTSS 789
 QY 303 RUTSHLGTVMVYLLSMLGTHDKDMSRTLLAMSSQDSCISNRQSGCPLLIQLLHG 362
 Db 790 DLPESH--DMLHNPYS--ALSSAGNIMPTTY-SVTFKI--G 824
 QY 363 NDKSVLLGNSRGSKEARASAAALHNIHS--OPDDKGRGRREIRVLHLEQIRAYCET 419
 Db 825 MKGPALTKRKFSGRPRSRQGWSTNTVSPFSWSPDISBGEIIFKPQLPG--SA 878
 QY 420 CMEQBAHEPGMDQDNMPAPVEHQ--ICPACVLMKLSFDEEHRH 464
 Db 879 INSIKVGSGFPGKRRPRGAGLSGRGRGKJASGIGAVLPGVSTADISSNKKDEEN 938
 QY 465 ANNELGGLQAIABELLQVDCMYGLTDHYSITLRYAGMALTNLTFGVANKATL--CSMK 523
 Db 939 SMHN--TVLFPSSDKFTLMD--MCVVGSGFGQAGRLLACQC 980
 QY 524 G-CMRALVAQLKSESDLOQVIASVLRNLWSR--ADVNSKKT-- 562
 Db 981 GQCYHFCYSIK--ITKVLSKGWRCLECTVCEACGKATDGRLLLCDDCDIS 1031
 QY 563 --LREYGVKALMECALEYKKESTLKSVLAL-- 592
 Db 1032 YHTYCLDPLQTVPKGKWKMCWCRHCGATSAGLRC--EWONNYTOCAPCASLSSCPV 1089
 QY 593 --WNL5AHC--TENKADI-- 606
 Db 1090 CYRNYREEDLILOCROCDRW--MHAVCONLTHEEVENVADIGDCMCRPYMPASNPSS 1148
 QY 607 -CAVDGALAPLVTGL--TYRQNTLAIIESGGGILRNVSLLATNEDHR-- 653
 Db 1149 DCESSILVAQIVTKVKELDPPKTYTO--DGCVLTESGMTQLQSLTVTPVRKRSPKPKL 1206
 QY 654 QILRENN--CLOTL--LOHLKSHSLTIVSNACCTLNLSARNPK--DOBALMD--MGAVSM 706
 Db 1207 KIINQNSVAVLQTPDIQSHSDRGEMDSREGELMDCDGKESPEREAVDEIKGVEG 1266
 QY 707 LKMLIHSKRWIANG--SAAALRNLMANRPAYKQANIMSGSLP5LHVYKQKALAEULD 765
 Db 1267 TDGVKKRKRKPYRPGIGGFVVRORSRTGQKTKRSVIRKDSG--SISEQLPCRDG 1322
 QY 766 AQLSETFDNIDNLSKASHRSKORHK--QSLXGVDVFTNRHD-- 807
 Db 1323 SEQLPDLTVDESVSVESTETEKIKRKRKKNKLEETFPAYLQBAFPGKDLTTSQSKIS 1382
 QY 808 -DNRSO--NFNTGCM--TVLSPYLNTTVLPSSSSSRGSLDSRSEKDRSLERE- 855
 Db 1383 LDNLSEDAQLLYKTNMTGFLPSLDPLLSSSSAFTKSGTHGPADDPLADISEVLNTDD 1442
 QY 856 --RGIGLGNYPATENPGT--SSKRL--QISTT 883
 Db 1443 DILGITSDDLAKSVSDHSDIGEVTDPSLPPQPNVQSSRPLSEBQLDGLSPELDKMTD 1502
 QY 884 AAOIAKYM--EVSIAHYS--QEDRSS-- 906
 Db 1503 GAILGKLYKIPELGKDVEDLFTAVLSPANTQTPPLPQPPPTQLPIHQDAFMRPLM 1562
 QY 907 --GSTTELHCVTDENALRRSSAAHTS--NTYNFKSE-NSNRTC 947
 Db 1563 NGLTSSPHL--PHNSLPPSGGLTFSALIAQSSYPDARDKNSAFNPASDPNNSWTS 1617
 QY 948 SMPYAKLEYKSSNDLSNVS--DGYGKRGOMKPSIESYEDD--ESKFC 996
 Db 1618 SAPTVEGE--NDTMSAQRSTLKWEKEALGEMATVAPVL--YTNFNPKEEFPD 1670
 QY 997 YGOYPADLAHKIHSANMD--DNDGELDTPINTYLSKYSEBQLNSGQSPSONE 1047
 Db 1671 WTRVKQIAKLWRKASSQERAPYVQKARDNRAAL--RIN-KVQMSNDSMKRQOQDSIDP 1727
 QY 1048 RWAEPKHILBDEIKQSEQ--FQSRNOSTTYPVYTEST--DDKHLKFQPH 1092

Db 1728 SSRIDSELPKPLQKRESEHEQWKFQROMROKSKQAKIEATQKLEQVKNFQOQOQOQ 1787
 QY 1093 FGOQECV--SPYRSGANGSETNRVGNHGINQNVQSLOCQ--DYVEDDKPNYS-- 1144
 Db 1788 FGSQHLVLQSGSDTPSSGIGSLTPQPGNG--NMSPAQSFHKLFTKQPPSTPTSTSSD 1844
 QY 1145 ERYSESEQHEEBEPNYSIKYNEEKHVDPDIDYLSKYATDIPSSQKOSRFSKSS-- 1202
 Db 1845 DVEFKPQAPPPAPPSRI--PQDSLQAOQTSQPPSPQVFGSGNSRP 1891
 QY 1203 --QSSKTEHSSSSENTSTPSSNAKRNQLHPSSAQSRSGOP 1243
 Db 1892 PSMYDPYAKMVGTPRPYPVGHFSRNSAPVEN--CTPLSSVSRLQMNATTANRPS--P 1948
 QY 1244 QKAATCKVSSINOTIQTVCVBTP--ICFSRCSLSSLSASADEICGNOT 1292
 Db 1949 VR-DLCSSSTTNNDPYAK--PPDTPRVMTDOPPKSLGLSR--SPVYSEQTAKGPAAAGTS 2004
 QY 1293 --TQEADSANTLOIAIEIKEKI--GTRSAEDPVSEVPAVSQSPRTKSSRL 1337
 Db 2005 DHFTKSPRADVFQORIPDSYARPLLTAPLDSGPGFKTQMPPPS--SQDPY-- 2057
 QY 1338 QGSLSESARHKAVFSSGAKSP--SKSGAQTPKSPPEHVYQETPLMFSRC 1387
 Db 2058 --GSVSQASRLSVDPYERDALTFRPIDNFHNSQNDPYSQPLTPEHPAVNES--FAHP 2112
 QY 1388 TSVSSLDSPFSRSLASVQSEPCGVMVGIISPSDLPSPGQTMPPSRSKTTP--PP-- 1442
 Db 2113 SRAFSQGTISRPTSQDPYQPPG--TPRPVDSYSSSGTARSNTDYSQPPGT 2165
 QY 1443 --PQATQKREVPKNKAPTAKRESGP-- 1467
 Db 2166 PRPTTVDPYQQOTPRPSTQTDLFTVETVNOQSHSDPYAHPPTGTPRGISVPYSQPPATP 2225
 QY 1468 --QBAVNAVAVQVVLPAADTLHFATE--STEDGFSCS-- 1503
 Db 2226 RPRISEGFTSSMTRPVLMEFNQDPFLQAAQNRGPAULPGPLVRPDP--TCSQTPRPPOGL 2283
 QY 1504 --SSLSALSLEDPFIQKDVLRIMPVQENONGNETE--SEQPK-- 1545
 Db 2284 SDTFSRVSPSAARDPYQSP--MTFRSQSDSFGTSQTAHDVADQPRPGSEGSFCASS 2338
 QY 1546 NENOEKBAEK--TIDSEKDL-LDSDDDDI--EILEECIISAMPTK 1586
 Db 2339 NSPMHSGQCGSVGSQPLPGVPVTSQVTDQTNVMAQADTEKLRQKREIILQOQOQK 2398
 QY 1587 S-SKAKKPAQATASKLPPPVARKPSQPLPVYKLLPSQNRLOPKHVSFT--PGDDMP 1639
 Db 2399 KIARGKSGSDSAVPHG--PLQHWQPVENVQAFTRPYPVPGNIRS 2445
 QY 1640 RVYCVETGPIFNSTATSLDLTIESPPNELAAG--EGVRGGAQSGEFKEDTIPTEGR- 1695
 Db 2446 PV--APPLGPYAVFPKQORGYPDPDASMGMRPHGFRFGPG--SHGTMPSQERF 2498
 QY 1696 --STDEAQQGKTS--SVTIPELDDNK--AEEGDIAEC- 1727
 Db 2499 LVPPQOQIGSGVSPQLRARSVDMPRLNNSQMNPNVGLPQHFSQSLPVQOHNILQOAY 2558
 QY 1728 --INSAMPKPKSHKPFVVKIMDQVOOASAS--SSAPKNQOLGKKKKPTS-- 1774
 Db 2559 IELRHRAFDGRQLPFPAPP--GSVVEASSNLRHGNFIPRPDPFGPRHTDPMRAPPQGLP 2616
 QY 1775 --PVKP--IFONTYRTRVRKNADSKNNIN-- 1800
 Db 2617 NQLPVHPDLEQVPPSQOEGSHSSVHSSVMVMTLNLHPJGGFSEAPLSTSPSETTSNLQ 2676
 QY 1801 --ABRVFSDNKKDKKONLKN--NSKDFNDK--LPNNEDVRGSPAFD 1841
 Db 2677 ITTQPSDGLBEKLDSDPSVKELDKOLEGVVEKLDDELDLENLNLDTEDGKVVELDTLD 2736
 QY 1842 SPHYTPIEGTFCFSRNDLSLSDFD--DDVDLSREKA--EL-- 1881
 Db 2737 NLENTNDP--NLDLLRSRGEFIIAYTDELMDGDKKSMFNEELDLPIDDKLNDQ 2788

CC -!- SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
 CC reorganization to the kinetochore/centromere (coronal surface of
 CC the outer plate) and the spindle during mitosis.
 CC -!- DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
 CC -!- PTM: Hyperphosphorylated during mitosis.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U19769; AAA82889.1; --
 CC EMBL; U30872; AAA82935.1; --
 CC EMBL; U25725; AAA86889.1; --
 CC PIR; PC4035; PC4035.
 CC Genew; HGNC:1857; CENPF.
 CC GK; P49454; --
 CC MIM; 600236; --
 CC GO; GO:0005699; C:kinetochore; TAS.
 CC GO; GO:0005634; C:nucleus; TAS.
 CC EMBL; U25725; AAA86889.1; --
 CC GO; GO:0005619; C:spindle; TAS.
 CC GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
 CC GO; GO:0007088; P:regulation of mitosis; TAS.
 CC Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
 CC Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat; Polymorphism;
 CC Lipoprotein; Prenylation.
 CC FT DOMAIN 14 197
 CC FT DOMAIN 273 769
 CC FT DOMAIN 823 1328
 CC FT DOMAIN 1642 1746
 CC FT DOMAIN 1862 2987
 CC FT DOMAIN 2207 2568
 CC FT REPEAT 2207 2386
 CC FT REPEAT 2389 2568
 CC FT DOMAIN 3015 3032
 CC FT LIPID 3207 3207
 CC FT VARIANT 3202 3202
 CC FT CONFLICT 16 16
 CC FT CONFLICT 250 250
 CC FT CONFLICT 272 272
 CC FT CONFLICT 611 611
 CC FT CONFLICT 1494 1589
 CC FT CONFLICT 1611 1611
 CC FT CONFLICT 1811 1811
 CC FT CONFLICT 2242 2243
 CC FT CONFLICT 2335 2335
 CC FT CONFLICT 2492 2492
 CC FT CONFLICT 2545 2561
 CC SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
 CC
 CC Query March 2.68; Score 379.5; DB 1; Length 3210;
 CC Best Local Similarity 17.94; Pred. No. 1.3e-06;
 CC Matches 574; Conservative 502; Mismatches 1243; Indels 895; Gaps 138;
 CC
 CC 7 DQLLKQVEALKWENSLRQELNDSNHLTKLETFASNMKEVLKQLOSGSIEDEAMSSGQI 66
 CC 475 NELRSMEEMKENLLKSHSEKAREVCHLEAEKLNKQCLNQSNFADKAKNTSQE 534
 CC 67 DLLRLKE-LNLDNSNTPGVKLRKMSLRYSYSGREGSVSRGECSPVPMGSPFRGFVN 125
 CC 535 TMLRDLOEKINQENSLTLEKL-KLAVADL-----EKQDCSDOLL----- 574
 CC 126 GSRESTGVLEL-----EKERSLLADLDKKEKMDWYAYQNLNLTKEIDSLPLTENF 178
 CC 575 --KKREHHIQNDKLSKTESKALLSALELKKE---YEELKE-EKTLFSCWKSENE 627
 CC 179 SLQDTMTTRQLEYEARQIRVAMESQLGTC-QDMKRAQRIARIQIQIEKD-----ILR 230

Db 628 KLLTQM---ESEKENLQSKI---NHLETCLTKQOIKSHNEYNERVRTLEMDRENLSVEIRN 681
 QY 231 IROLLQSOATEARSS-----QNKHETGSHDAERQNEGQGVGEINMATSNGQGGTTRM 284
 Db 682 LHNVLDSKSVETQKLYANMELQOKAEFSQKQKEIE-----NWCLK----- 724
 QY 285 DHETASVLSSTSTHAPRRLTSHLTKVEMVYLLSMLGTH--DKDDMGRTLLAMSSQD 342
 Db 725 -----TSQLTGVDELEHKLQLLSNEIMDKRCYQDLHAYES-- 762
 QY 343 SCISMROSGLPLLIQLHNGNDKSVLLGNSRSGKEARA-RASAALH-----NIHSQPD 397
 Db 763 -----LRDLL--KSDASLVTHEDHQSLLAFDQPPAMHSHFANIIIGQGM 807
 QY 398 KRGRREIRVLHLEQIRAYCETCWEQEAHEPGMDQKPFM-PAPVEHQICPAVCVLMKL 456
 Db 808 PSESECR-----LEADQSPKNSAILQNRV-----DSL 835
 QY 457 SFDEHRHAMNELGGLQAIABLLQVDCMTGLTNDHYSITLRYAGWALTNLTFGDVANK 516
 Db 836 EFSLESQKQMN-----SDLQKQCE-----ELVQIKGEIEENLMKABQMHO 875
 QY 517 ATLCSMKGCVRAVLAQKSEEDLQOVIAVSLRNLSWRADVNSKKTILREVGSKVLMMECA 576
 Db 876 SFVAETS-----QRISKLEQEDTSAHONVVAETLSALE-----NKEKELQLLNDKVEIEQAE 926
 QY 577 LEVKKEST--LKSVLSALWNLNSAHTENKADI CAVDGALAFVLGTLTYRSQNTLAIIES 634
 Db 927 IQELKXNHLLEDLSKELQLLS-----ETLSLEK- 955
 QY 635 CGGILRVSSLIATNE-DHRQILRENNCLQTLLOHKLKSHSLTIVSNACGTLWNLARNPK 693
 Db 956 -----KEMSSIISLNKREIELTQENGTLKEINASLNQEKWNLQKS--ESFANYIDEREK 1009
 QY 694 DQEAWDMGAVSMKLNLIHSHKMIAMGSAALRNLMANPAKYKADANIMSPGSSLSLH 753
 Db 1010 SISELSDQ-----YKQEKLLILQRCBEETGNAYEDLSQYKAAQ----- 1047
 QY 754 VRKQALAEALDAQLHSETFDNINLSPKASHRSKQHKQSLGYDYFDINRHDNRSN 813
 Db 1048 -EKNSKLECLLN--ECTSLCENKRNLEQLKEAPAKHQEFL----- 1086
 QY 814 FNTGNMTVLSPLYNTTVLPSSSSRSGLSRSSEKDRSLERERIGLGNHYHPATENPGTS 873
 Db 1087 -----TKLAFAEERQNLMELETVQOALRSEMTDNQNSKSEAG----- 1126
 QY 874 SKRGL--QISTAAQIAKMEVSAIHQSQDRSSGSTTELHCVTDERNALRSSAAHTH 931
 Db 1127 --GLKQEIIMTLKBEQNKQKQVNDLLOENBQLMKVMTKHECQNLSEPIRNS----- 1177
 QY 932 SNTNYFTKSENRTCSM-PYAKLEYKSSNDLSNS-VSSSDGYGKRGKQMPKSTESYSED 989
 Db 1178 -----VKRESERNQCNPKQMDLEVEKISLSYNAQLVQLEAMLRNKELK-----LQES 1227
 QY 990 DESKFCYGOYPADLAHKHS-----ANHMD-----NDGELDTPIYNSLYKYSDE 1034
 Db 1228 EKEKEC-----LQHELQITRGDLETSLNQMQSOBISGLKDCED-----AEE 1270
 QY 1035 QLNSG--RQSPQNERWARPKHI-----IEDETKQSEQRQSRNQSTTYPVYVTESTDDKH 1086
 Db 1271 KYISGFHELSTQNDN-----AHLQCSLQTTWKNKLEKICEILQAEKYLVLTLNLSR- 1325
 QY 1087 LKFPHQFQQQCVCPYRSGANGSETNRVGNHGINQVNSQSLCOEDDYEDDKTNTYSE 1146
 Db 1326 -----SECITATRK-----MABEVGKLINLVKILNDDSGLLHGL 1360
 QY 1147 YSEERQHEERPTNYISKYNEEKHVDPIDYSLKYATDIPSSQKQSFPSK--SSSQS 1205
 Db 1361 VEDIPGGFQGGP-----NEQHPVSLAPLDENSGVEHLTSLDKVEQMFAELQKFLS 1413
 QY 1206 SKTEH-----MSSSENTSTPSSNAKRNQHPSSAQSRSGQ----- 1242
 Db 1414 LQSEKILHDQHCQSSKQVSELQTVVDSLKAENLVSLTNLNFQCDLVKEMQLGLEGLV 1473

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:05:26 ; Search time 183.5 Seconds
(without alignments)

4888.383 Million cell updates/sec

Title: US-09-442-489F-7

Perfect score: 14566

Sequence: 1 MAASVYQLLKQVLAALXNEN.....ESSCTQSPKRGSGYLTVTSV 2843

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_todent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10544	72.4	2829	13 P70039	P70039 xenopus lae
2	5162	35.4	1056	11 Q8BNP7	Q8bnp7 mus musculus
3	3512.5	24.1	2274	11 Q9ZIK7	Q9z1k7 mus musculus
4	3415.5	23.4	2303	4 Q95996	Q95996 homo sapien
5	2232	15.3	489	11 Q8BRD8	Q8brd8 mus musculus
6	2170.5	14.9	733	4 Q9UEZ1	Q9uezi homo sapien
7	1991.5	13.7	1246	4 Q9Y632	Q9y632 homo sapien
8	1780.5	12.2	2416	5 P91667	P91667 drosophila
9	1765	12.1	2417	5 Q9VAS9	Q9vas9 drosophila
10	1705	11.7	1685	4 Q9UEM8	Q9uem8 homo sapien
11	1252	8.6	324	11 Q8C493	Q8c493 mus musculus
12	1222.5	8.4	1067	5 Q9Y1T2	Q9y1t2 drosophila
13	1221.5	8.4	1067	5 Q961B0	Q961b0 drosophila
14	1041	7.1	208	4 Q9P119	Q9p119 homo sapien
15	756	5.2	159	4 Q7Z2Q8	Q7z2q8 homo sapien
16	680	4.7	146	11 Q8C919	Q8c919 mus musculus

17	660.5	4.5	16223	5	Q8IR22	Q8ir22 drosophila
18	649	4.5	5327	5	Q76891	Q76891 drosophila
19	647.5	4.4	5412	5	Q9W596	Q9w596 drosophila
20	564.5	3.9	2310	16	Q8CMU7	Q8cmu7 staphylococ
21	554	3.8	2296	4	Q9UHA8	Q9uha8 homo sapien
22	543	3.7	2607	11	Q8BTI8	Q8bti8 mus musculu
23	541	3.7	108	4	Q9HAM6	Q9haw6 homo sapien
24	537	3.7	2768	5	Q9VC00	Q9vc00 drosophila
25	528.5	3.6	2283	2	Q8VQ99	Q8vq99 staphylococ
26	527.5	3.6	3257	4	Q9V736	Q9v736 drosophila
27	526	3.6	2752	4	Q9UQ35	Q9ug35 homo sapien
28	517	3.5	2271	16	Q99QY4	Q99qy4 staphylococ
29	515.5	3.5	2275	16	Q8NUJ3	Q8nuj3 staphylococ
30	513	3.5	2232	5	Q8IFX6	Q8ifx6 caenorhabdi
31	511.5	3.5	1186	5	Q21227	Q21227 caenorhabdi
32	510	3.5	3392	3	Q7ZA38	Q7za38 ashbya goss
33	510	3.5	4322	10	Q7XXN1	Q7xxn1 oryza sativ
34	509.5	3.5	1188	5	O62302	O62302 caenorhabdi
35	509	3.5	3381	2	Q9KX33	Q9kx33 streptococ
36	505	3.5	18519	5	Q8ISF6	Q8isf6 caenorhabdi
37	505	3.5	18534	5	Q8ISF7	Q8isf7 caenorhabdi
38	498.5	3.4	7210	5	Q9V7G8	Q9v7g8 drosophila
39	498.5	3.4	9270	5	Q8MLD9	Q8mid9 drosophila
40	492	3.4	3111	5	Q9VH10	Q9vh10 drosophila
41	489	3.4	3072	2	Q939N5	Q939n5 streptococ
42	486	3.3	2178	2	Q9KWR3	Q9kwr3 streptococ
43	483.5	3.3	3443	11	Q8JZM8	Q8jzm8 mus musculu
44	483	3.3	2910	10	Q9FND5	Q9fnd5 arabidopsis
45	477.5	3.3	3241	5	Q7YZH1	Q7yzh1 drosophila

ALIGNMENTS

RESULT 1

ID	P70039	PRELIMINARY;	PRT;	2829	AA.
AC	P70039;	P79934;			
DT	01-FEB-1997	(T-EMBLrel. 02, Created)			
DT	01-MAY-1997	(T-EMBLrel. 03, Last sequence update)			
DT	01-OCT-2003	(T-EMBLrel. 25, Last annotation update)			
DE	Adenomatous polyposis coli.				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Vlemnickx K., Wong E., Guger K., Gumbiner B.M.;				
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; U6442; AAB41671.1; -				
DR	InterPro; IPR008938; ARM.				
DR	InterPro; IPR000225; Armadillo.				
DR	Pfam; PF00514; Armadillo_seg; 4.				
DR	SMART; SM00185; ARM; 5.				
SQ	SEQUENCE 2829 AA; 310878 MW; 8A2BABDB770E496 CRC64;				

Query Match 72.4%; Score 10544; DB 13; Length 2829;
Best Local Similarity 74.4%; Pred. No. 0;
Matches 2131; Conservative 263; Mismatches 413; Indels 58; Gaps 40;

Qy	1	MAAASVDQLLKQVEALKMENSLRQBLEDSNHLTKLETEASNMKEVLKQLQGSIDEAM	60
Db	1	MAAASVDQLKQVEALTKLETEASNMKEVLKQLQGSIDEAM	60
Qy	61	ASSGQDILLERLKLNLDSNPFVKLRGKMSLRSYGSRGCSVPVPMGSPFR	120
Db	61	ASSGPDILLERFKDLNLDSNIPAGKARPKMSRYSYSGRGLSGHSGCSVPVGSFQR	120
Qy	121	RGFVNGSRSTGYLELEKERSLLLDLKKKEKKWYVAQLQNLTKRIDSPLTNFSL	180
Db	121	RGLLNGSRGAGYWELEKELLLIAHEKEKEKWEYVAQLQNLTKRIDSPLTNFSM	180

181 QTMTRQLEBYEARQIRVAMEBOLGTCQDMKEKQARRIARIOIEKDIIRIOLQOSQAT 240
 181 QTMTRQLEBYEARQIRVAMEBOLGTCQDMKEKQARRIARIOIEKDIIRIOLQOSQVA 240
 241 E-AERSSONKHETGSHDABRONQGVGEINMATS--GNGGSGTTRMDHETASVLSSTH 298
 241 EAEERTPQSHDAGSRAEKLPGQQTSEITASCNVGSGQSSRADHDTTVMGSENSTY 300
 299 SAPRLTSHLGTVMVYSLLSMLGTHDKDDMSRTILLAMSSSQDSCISMQRGCGPLLIQ 358
 301 SVPRRLTSHLGTVMVYSLLSMLGTHDKDDMSRTILLAMSSSQDSCIAMRQSGCGPLLIQ 360
 359 LLHGNKDSVLLGNSGSKSEAPASAAALHNIHSDPDOKRGREIRVLHLEQIRAYCE 418
 361 LLHGNKDSVLLGNSGSKSEAPASAAALHNIHSDPDOKRGREIRVLHLEQIRAYCE 420
 419 TCWEQEAHEPFGDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIAB 478
 421 TCWEQEAHEPFGDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIAB 480
 479 LQVDCMYGILTNDHYSITLRYAGMALTNLTPGDVANKATLCSMKCMALVAQLKSESE 538
 481 LQVDCMYGILTNDHYSITLRYAGMALTNLTPGDVANKATLCSMKCMALVAQLKSESE 540
 539 LQOQVIASVLRLNLSWRADVNSKKTLEVGSVKALMECALEVKESTLKSVLGALNLSAH 598
 541 LQOQVIASVLRLNLSWRADVNSKKTLEVGSVKALMECALEVKESTLKSVLGALNLSAH 600
 599 CTENKADICAVDQALAFVLTLYRSQTNLTALIESGGGILRNVSLLIATNEDHQILRE 658
 601 CTENKADICAVDQALAFVLTLYRSQTNLTALIESGGGILRNVSLLIATNEDHQILRE 660
 659 NNCLOTILQHLKSHSLTIIVSNACGLTNLSARPKDOEALWDMGAVMLKNIHSHKMI 718
 661 NNCLOTILQHLKSHSLTIIVSNACGLTNLSARPKDOEALWDMGAVMLKNIHSHKMI 720
 719 AMGSAALRNLMANRPAKYDANIMSPGSLPSLHVRCOKALEAEADAHLSETFDNIN 778
 721 AMGSAALRNLMANRPAKYDANIMSPGSLPSLHVRCOKALEAEADAHLSETFDNIN 780
 779 LSPKASHESKORHOSIYDGVDFDTHRDDN--RSDNENTGNMTVLSPLYNTVLPSSSS 836
 781 LSPKTHRNKORHOKMLCSYALDSSRHDDICRDSIGNLTVLUSPYNITVLPSSSS 840
 837 SRGSLDSSREKDRSLERERIGLGNHYPATENPGTSSKR--GLQISITAAQIAKYMEEVS 895
 841 PRPTMDGSRPEKD---RERTAGLGNHYSITTESSGNSKRIQIQLSTT-AQISKVMEVS 895
 896 AIHTSQBDRSSGTTTELHCVTDERNALRPSAAHNTNFTKSNENRNTCSMPYAKLE 955
 896 NIHLVQENRSSGASAEHCHMSDERNSORPSSNHPQSNPFTFKAESSTRGCGPVAFMKE 955
 956 YKRSSNDSLNSVSSDGYGRGOMKPSIESYDDDESKYGOVPADLAHKIHGANHMD 1015
 956 YKVASNDSLNSVESTGYGKGVKPSVESYEDDESKEFYSYGOVPAGLAHKIQSANHMD 1015
 1016 DNGELDTPINYSKYSDQOLNGROSQONERWAPKHIIEDEIKOSQOROSNOSTY 1075
 1016 DNDTELDTIPINYSKYSDQOLNGROSQONERWAPKHIIDSEMKQSQORQPTTKTY 1075
 1076 PVYTESDTHLQFOPHFGQECVSPY-RSRGANGS-ETNRYVSGHGINOVYSQSLCOED 1133
 1076 SSTENKEEKHKKFPHPFNQSENVPAYTRSGANNQVDQSRVSSNLSKASKHCHQVD 1135
 1134 DYEDDPTNYSERYSEEEQHEEF-ERPTNYSIK-YNEEKXHVDPDIDYSLKYATIPSS- 1190
 1136 DYDDDKTNPISERYSEEEQHEEF-ERPTNYSIK-YNEEKXHVDPDIDYSLKYATIPSS- 1195
 1191 OKOSFSPKSSGQSSKTHMSSESTSTPSSNAKRONLHPSSAQSSG--QPKAAT 1248
 1196 QKPSFPYSNNSSKQPKKEQVSSNS--NTPTPSPNSNRQNLHPNSAQSPGLNRPAQIPN 1254

1249 CKVSSINOETIQTVCVEDTPICFSSCSLSLSAEDEI--GCNQTTOEADSANTLOIAEI 1307
 1255 -KPSINOETIQTVCVEDTPICFSSCSLSLSAEDEIIEGRNRSQGSSNTLQITTEP 1313
 1308 KEKICTASADPVEPAVQHPRTKSSRLQGSLS--SESARHKAVEFSSGAKSPKSGA 1366
 1314 KE-ISAISKOGAVNETRSSVHTTKNRLQTSNISPSDSSRHKSVEFSSGAKSPKSGA 1372
 1367 QTPKSPPEHYVQETPLMFSCSTSVSSLDSFESSIASVSEPCSS--GMVSGIISPDLDP 1425
 1373 QTPKSPPEHYVQETPLMFSCSTSVSSLDSFESSIASVSEPCSS--GMVSGIISPDLDP 1432
 1426 SPGOTMPPSRKSTPPPPQTAQKREVPKNAPTAEKRESGPKQAAVNAAVQVQVLPDA 1485
 1433 SPGOTMPPSRKST--PPPPQTAQKREVPKNAPTAEKRESGPKQAAVNAAVQVQVLPDA 1489
 1486 DTLHFAFTESTPDGFCSSSLSALSDELDEPFIQKDVLRIMPVPOENDNGNETSEBPKE 1545
 1490 DTLHFAFTESTPDGFCSSSLSALSDELDEPFIQKDVLRIMPVPOENDNGNETSEBPKE 1547
 1546 NENOEKAEKTIIDSEKOLLDDDDDEIIEBECIISAMPTKSKRKA--PAQTASKLPPP 1604
 1548 IDNKAKEKRESEKEDMLDDT-DDDIDIEBECIISAMPTKSKRKA--PAQTASKLPPP 1606
 1605 VARKPSOLPVYKLLPSQNRLOPKQVSYFTPGDDMPRYVCVEGTPINPSTATSLDTIES 1664
 1607 VARKPSOLPVYKLLPSQNRLOPKQVSYFTPGDDMPRYVCVEGTPINPSTATSLDTIES 1666
 1665 PPNELAAGEVGRGAOSGEPEKDTIETEGRSSTDEACGGKTSSTIPELDDNKAEGDIL 1724
 1667 PPSE-PTNDQONTDSLTDEKDTIETEGRSSTDEACGGKTSSTIPELDDNKAEGDIL 1724
 1725 AECINSAMPKSKHPRVVKIMDOVOQASASSAPKNOL--DGKKKKTPSPVKPIQNT 1783
 1725 AECIHSAMPKSKHPRVVKIMDOVOQASASSAPKNOL--DGKKKKTPSPVKPIQNT 1784
 1784 BYRTRVRKADSKNVLNABRVFSDNKSQKQNLKNSKDFNDKLPNNEDVRGSPAFDSP 1843
 1785 GFKEKRLKNTLKNLPSNENQYC---DPRKPSKSKPKNVANEKIPNNEETKGF--FAPDSP 1840
 1844 HVTPIEGTGYCFERNDSLSDPDVDSREKAEKELKAKENKSEAKVTSHETLSN 1903
 1841 HVTPIEGTGYCFERNDSLSDPDVDSREKAEKELKAKENKSEAKVTSHETLSN 1900
 1904 QOSANKTOATAKPIRGOKPILQKOSTPQSSKIDIPDRGAATDEKLQNFALNTPVCF 1963
 1901 NPMGKQDQTPKSLGGRDQPKALVQKPTSFSAAGTQDRGGATDEKMFALNTPVCF 1960
 1964 SHNSLSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSPHVEDTPVCFSR 2023
 1961 SRNSLSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSPHVEDTPVCFSR 2020
 2024 NSLSLSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSPHVEDTPVCFSR 2081
 2021 NSLSLSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSPHVEDTPVCFSR 2080
 2082 DIQPDSEHGLSPDSENFDMKATOEAGANSIVSSIHQAAAACLSRQASSSDSLSLKSG 2141
 2081 DIQPDSEHGLSPDSENFDMKATOEAGANSIVSSIHQAAAACLSRQASSSDSLSLKSG 2140
 2142 ISLSPHPLTPDQBEKEFTSNKGPRIILKPGKSTLETTKIESKGIKGGKVKYKSLITG 2201
 2141 ISLSPHPLTPDQBEKEFTSNKGPRIILKPGKSTLETTKIESKGIKGGKVKYKSLITG 2200
 2202 KVRNSISISGOMKQPLQANMPSISRGTMTHIICVRNRSSTSPVSKGKPLTPASKSP 2261
 2201 KSRSSDPSFCHCKQSVQTNMPSISRGTMTHIICVRNRSSTSPVSKGKPLTPASKSP 2260
 2262 SEGOTATTSRGAKPSVKSELSPVARQTSQIGASSKAPSRGSRDSTPSPQAQPSRPI 2321
 2261 NENFSSSSPKGTIP-LKSELVYGSRSSTPGGSSKNSRSGSRDSTPSPQAQPSRPI 2319
 2322 QSPGRNISISGRNIGISPPNKLSQLPRTSPSTASTSSGSKKMSYTSPPGMSQONLTKQ 2381

Db	2320	QSPGRNSISPGKNGISPPNFKSQLPRTTSPASTKSSGSRMSYTSFGRLOSPNLSKQ	2379
Qy	2382	TGLSKVAGSIPRSESASGLNNGANKVELSRMSSTKSSGESDRGPRVLRQS	2441
Db	2380	SGLPKTHSIPRSESASGLNQ-VNWTGSKVELSRMSSTKSSGESDRGERPALVRQS	2438
Qy	2442	TPIKAPPTLRKLEESASPELSIPSRAPSPTRSQATQVLPSPDMSLSTHSSVQA	2501
Db	2439	TPIKAPPTLRKLEESASPELSIPSRADSPRSPQTQ2ALSPSLPDMSLSTH-SIQ	2497
Qy	2502	GGWRKLPNLSPTIYNDGRAPKADHARSSESPLNRSCTGWTKEHSHSSSLPRV	2561
Db	2498	GGWRKLPNLSPTIYNDGRAPKADHARSSESPLNRSCTGWTKEHSHSSSLPRV	2555
Qy	2562	STWRITGSSSSILSASSSESSEKAKSEDEK-HVNSISGTQKSKENQVSAKQWRIKENE	2620
Db	2556	STWRITGSSSSILSASSSESSEKAKSEDEKQVCSPPGPR--SECSSSAKQWRIKENE	2613
Qy	2621	--SPNISTQVSSGATNGASKTLYOMAPAVSKTEDVWRIEDCPINPRSGRPTGN	2678
Db	2614	LETSPNGSSSTIAE-SNCSLESKTLVYQMAPAVSKTEDVWRIEDCPINPRSGRPTGN	2672
Qy	2679	TPPVIDSYSSEKAMPNIKDKONQAKONGSVPMRTVGLNRLNSFIQVDAPQKGTET	2738
Db	2673	SPPVLDNVDGQKE-EAAKCHTEHNSGNGVPL---LENROKSFIVDGLDTKGTDP	2727
Qy	2739	KPGQNPVPSVETNESSIVERTPPSSSSSSSKHSPSGTVAARVTFNNPSPRKSADST	2798
Db	2728	KSLINNQ---OFTNENTVAERTAFSSSSSSSKHSPSGTVAARVTFNNPSPRKSNGENS	2784
Qy	2799	SARPSQIPTPVNNNTKRDSTDSPESSGTOSPKRHSGSYLVTSV	2843
Db	2795	TSRPSQIPTPVNTSKRDSTETTTDSSGSPKRHSGSYLVTSV	2829

RESULT 2

QBNP7	QBNP7	PRELIMINARY;	PRT; 1056 AA.
AC	QBNP7		
DI	01-WAR-2003 (TRENBLrel. 23, Created)		
DT	01-WAR-2003 (TRENBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Adenomatosis polyposis coli (Fragment).		
GN	APC.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Adipose tissue;		
RX	MEDLINE=22354683; PubMed=12466851;		
RA	The FANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs."		
RL	Nature 420:563-573(2002).		
DR	EMBL; AK080907; BAC38073.1; --		
DR	MGI; MGI:88039; APC.		
DR	GO; GO:0005737; C:cytoplasm; IDA.		
DR	GO; GO:0005634; C:nucleus; IDA.		
DR	GO; GO:0008013; F:beta-catenin binding; IDA.		
DR	GO; GO:0009952; P:anterior/posterior pattern formation; IMP.		
DR	GO; GO:0009798; P:axis specification; IMP.		
DR	GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.		
DR	GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000225; Armadillo.		
DR	Fam; PF00514; Armadillo_seg; 4.		
DR	PSIT; PS50176; ARM_REPEAT; 1.		
FT	NON TER 1056		
SQ	SEQUENCE 1056 AA; 117633 MW; 86C79FDC12C23FAB CRC64;		

Qy	1	MAAASDQLLKQVEALKMENSNRQELNDSNHLTKLETEASNKVKVLKOLQSGIDEAM	60
Db	1	MAAASDQLLKQVEALKMENSNRQELNDSNHLTKLETEASNKVKVLKOLQSGIDEAM	60
Qy	61	ASSGQIDLLRLKELNLDSSNPFQVKLRKMSVRSYSGSREGSVSSRGSGSPVPMGSPFR	120
Db	61	-TSQQIDLLRLKELNLD-SNPFQVKLRKMSVRSYSGSREGSVSSRGSGSPVPMGSPFR	118
Qy	121	RGFVNGSRESTGYLEBELEKERSLLADLDKEEKEKDWYQAQLQNLTKRIDSLETFENFSL	180
Db	119	RTEFVNGSRESTGYLEBELEKERSLLADLDKEEKEKDWYQAQLQNLTKRIDSLETFENFSL	178
Qy	181	QDTMTROQLEVEARQIVAMEEQLGTCQDWEKPAQRARIQOIEVDILRIROLLOSOAT	240
Db	179	QDTMTROQLEVEARQIVAMEEQLGTCQDWEKPAQRARIQOIEVDILRIROLLOSOAA	238
Qy	241	EAERSSQNHETGSHDAERQNGQGVGEINMATSNGQGSTTRMDHETASVSSSSSTHSA	300
Db	239	EAERSSQSRHDAASHEAGRQHEGHAESNTAASSSGSQSPATFVDHETASVSSSGTHSA	298
Qy	301	PRELTSGLTKVEMVYSLLSMLGTHDKDMSRTLLAMSSQSDSCISMRQSGCLPLLIQLL	360
Db	299	PRELTSGLTKVEMVYSLLSMLGTHDKDMSRTLLAMSSQSDSCISMRQSGCLPLLIQLL	358
Qy	361	HGNDKDSVLLGNRSGKEARASAAAHNIHSQPDQKGRREIRVHLHLEQIRAYCETC	420
Db	359	HGNDKDSVLLGNRSGKEARASAAAHNIHSQPDQKGRREIRVHLHLEQIRAYCETC	418
Qy	421	WEQEAHEPQMDQKMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIABELQ	480
Db	419	WEQEAHEPQMDQKMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIABELQ	478
Qy	481	VDCEMYGLTNDHYSITLRRVAGMALTNLTFDGVANKATLCSMKGCWRALVAQLKSSIDL	540
Db	479	VDCEMYGLTNDHYSITLRRVAGMALTNLTFDGVANKATLCSMKGCWRALVAQLKSSIDL	538
Qy	541	QOVIASVLRNLSWRADVNSKKTLYRQVSKALMECALEVKESTLKSLSALWNLSAHCT	600
Db	539	QOVIASVLRNLSWRADVNSKKTLYRQVSKALMECALEVKESTLKSLSALWNLSAHCT	598
Qy	601	ENKADICAVDGAFLVGLTLYRSQNTNLAIIESGGIILRVSSSLIATNEDHQILRENN	660
Db	599	ENKADICAVDGAFLVGLTLYRSQNTNLAIIESGGIILRVSSSLIATNEDHQILRENN	658
Qy	661	CLQTLQLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSMKLIHSHKHMIAM	720
Db	659	CLQTLQLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSMKLIHSHKHMIAM	718
Qy	721	GSAALRNLMANRPARYKOANIMSPGSSLPFLHVRKQKALEAELDAQHLSSETPDNDLS	780
Db	719	GSAALRNLMANRPARYKOANIMSPGSSLPFLHVRKQKALEAELDAQHLSSETPDNDLS	778
Qy	781	PKASHESKORHKOSLYGDYVDFNRRHDDNRSDNFNTGNMTVLSPYLNTTLPSSSSSRGS	840
Db	779	PKASHESKORHKONLYGDIYAFDNRHDDNRSDNFNTGNMTVLSPYLNTTLPSSSSSRGS	838
Qy	841	LDSSRSEKORSERERIGLGNYPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS	900
Db	839	LDSSRSEKORSERERIGLGNYPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS	898
Qy	901	QEDRSSGSTTELCVTDENALRRSSAATHNTYNTKSENSENRTCSMPYAKLEYKRS	960
Db	899	QEDRSSGASTTEPCVADDSARRSSASHTHNTYNTKSENSENRTCSMPYAKLEYKRS	958
Qy	961	NDSLSNVSSSDGYGKQKQMPKPSIESYSEDESKFCYGOYPADLAHKIHSANHMDNDGE	1020
Db	959	NDSLSNVSSSDGYGKQKQMPKPSIESYSEDESKFCYGOYPADLAHKIHSANHMDNDGE	1018

QY 1021 LDTPINYSKYDEQLNSGRQSPSONERWARPKHIED 1058
 DB 1019 LDTPINYSKYDEQLNSGRQSPSONERWARPKHIED 1056

RESULT 3
 Q921K7
 ID Q921K7 PRELIMINARY; PRT; 2274 AA.
 AC Q921K7
 DT 01-MAY-1999 (T-EMBLrel. 10, Created)
 DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE APC2 protein.
 GN APC2
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99147086; PubMed=10021369;
 RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
 RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.,
 RT "Identification of APC2, a homologue of the adenomatous polyposis coli
 RT tumour suppressor";
 RL Curr. Biol. 9:105-108 (1999).
 DR EMBL; AJ130783; CAAL0207.1; JOINED.
 DR EMBL; AJ130784; CAAL0207.1; JOINED.
 DR EMBL; AJ130785; CAAL0207.1; JOINED.
 DR EMBL; AJ130786; CAAL0207.1; JOINED.
 DR EMBL; AJ130787; CAAL0207.1; JOINED.
 DR EMBL; AJ130788; CAAL0207.1; JOINED.
 DR EMBL; AJ130789; CAAL0207.1; JOINED.
 DR EMBL; AJ130790; CAAL0207.1; JOINED.
 DR EMBL; AJ130791; CAAL0207.1; JOINED.
 DR EMBL; AJ130792; CAAL0207.1; JOINED.
 DR EMBL; AJ130793; CAAL0207.1; JOINED.
 DR EMBL; AJ130794; CAAL0207.1; JOINED.
 DR EMBL; AJ130795; CAAL0207.1; JOINED.
 DR EMBL; AJ130796; CAAL0207.1; JOINED.
 DR PIR; T30258; T30258.
 DR MGI; MGI.1346052; R0C2.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg_3.
 DR SMART; SM00185; ARM; 5.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
 SQ SEQUENCE 2274 AA; 243137 MW; 75ABDA15D0F707F5 CRC64;

Query Match 24.1%; Score 3512.5; DB 11; Length 2274;
 Best Local Similarity 34.3%; Pred. No. 4.9e-164;
 Matches 1005; Conservative 354; Mismatches 788; Indels 779; Gaps 97;

QY 4 ASVDQLKQVEALKWENSLRQELNDSNHLTKLETSAGNKEVLKQLOQSIIDEA--MA 61
 DB 6 ASVEQLVRQVEALKWENSLRQELNDSNHLTKLETSAGNKEVLKQLOQSEARVLV 65

QY 62 SSGQIDLLERLKEINLDSNFPVKYRSKVSLSYSGRSVSRGSCSPVPMGSPFR 121
 DB 66 SSGQTEVLEQLKALQTDISSLYNLKFPAP---ALGPEP---AARTPEGSPV-HGSGPSK 117

QY 122 -GFVNGRSTGYLEELKEKRSLLADLKEEKDWYQAQLQNLKRIQDLSPLTENFSL 180
 DB 118 DSFGELSRATIRLEEDQRCFLLSIEKEKELWYSLQGLSKRLELPHVDTFSM 177

QY 181 QTDMTREQLYEARKIRVAMEQLGTQCDMEKPAQRRIAFIQIEXDIL----RIQ--- 233
 DB 178 QMDLIRQLEFEAGHRSLSMEERFGTSDVMVQRAQIRASRLQIDKELLEAQDRVQOTEP 237

QY 234 --LLQSQATERSRSQNKHETGSHDAERQEGQGVGEINMATSGNGSGSTTRMDHETASV 291
 DB 238 QALLAVKPVAVEEKEAEVTPHEDGTPQ-----PCN----- 269

QY 292 LSSSSTHSAPRLTSHLTGTVEMVYLLMLGTHDKDDMSRTLLAMSSSDSCISMRQSG 351
 DB 270 -----SKVEVFWLLSMLATDOEDTARTLLAMSSSPESCVAMRRSG 311

QY 352 CLPLLIQLLHNDKDSV---LLGNRSRSGKEARASAAALHNIHSDOPDDKRGRRRLVHL 408
 DB 312 CLPLLIQLLHNDKDSV---LLGNRSRSGKEARASAAALHNIHSDOPDDKRGRRRLVHL 371

QY 409 LLEQIRAYCETCEWQSAHEPGMDQKNPMPAPVEHOICPAVCVLMKLSFDEHRAJME 468
 DB 372 VLEQIRAYCETCEWQSAHEPGMDQKNPMPAPVEHOICPAVCVLMKLSFDEHRAJME 426

QY 469 LGGLQIAELQLQVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKMGOMRA 528
 DB 427 LGGLQIAELQLQVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKMGOMRA 486

QY 529 LVAQLKSESDLOQVIAVSLNLSWRADVNSKTLREVGSGVKALMECALVKKESTLKS 588
 DB 487 IVAQLGSESELEHQVSSILNLSWRADVNSKTLREVGSGVKALMECALVKKESTLKS 546

QY 589 LSALWNLSAHCTENKADICAVDGAFLVGLTLYRSQTNLTALIESGGGILRNVSLLIAT 648
 DB 547 LSALWNLSAHCTENKADICAVDGAFLVGLTLYRSQTNLTALIESGGGILRNVSLLIAT 606

QY 649 NEDHQRLRENCLQTLLOHLKSHSLTIVSNAGCTLWLSARNPKQDEALWDMGAYMLX 708
 DB 607 REDYRQVLRDHNCLQTLLOHLKSHSLTIVSNAGCTLWLSARNPKQDEALWDMGAYMLX 666

QY 709 NLIHSHKMTAMGSAALRNLMANRPKAKYDANT-MSPGSSLSPLSHVRKQKALEAEADAQ 767
 DB 667 NLVESKHKMTAMGSAALRNLMANRPKAKYDANT-MSPGSSLSPLSHVRKQKALEAEADAQ 726

QY 768 HLSETFDNIDNLS-PKASHRSKQ-----RHQKSLYGDYVFTDTHRDDNRSNFTGNT- 820
 DB 727 HLVAHGLHLEKQSLPEAETTSKKPLPLRLHLDGLVQVYASDSGCFDDDDAFSAAAAATTA 786

QY 821 -----VLSPLYNTVLPSSSSRSGSLDSRSEKDRSLERERGLGLNHYHPATENPTSS 874
 DB 787 EPASPAVNSMFLPGFFLQGGALAR-----TPPARQGGLEAKAG-----GEAAVAA 833

QY 875 KRGLQISTTAQIAKAMEEVSIAHTSQEDRSSGTTTELHCVTDERNALRRSSAAHTSNT 934
 DB 834 KAKALALAVARIDRLVEDISALHTSDSDGFSLS-----GDPQAPREGRAQSCSPC 887

QY 935 YNFT--KSENNRTPCMPYAKLEYKRSNDLSNLSVSSDYGKGGKQMKPSIESYSEDES 992
 DB 888 RGTEGGRREGSRAHPLRLKAAHTSLNDSLSNGSGTSDGYCTREHMT- 936

QY 993 KFCYGVQVPADLAHKAHSAHMDNDGELDTPINYSKYSDEQLNSGRQSPSONERWARP 1052
 DB 937 --CLPAL-----AEHRDD-----PVRQ--TRP 956

QY 1053 KHIDEIKQEQEQSQSNQSTTPYVTESTDDKHLKQHPGQEQECVSPRSRANGSET 1112
 DB 957 RRLDILPSRAELPARDATATDARVT-----IKLSPTYQHVPFLD-----GAAGA-- 1002

QY 1113 NRVSNGHINQVNSQLCEDDYDDKPTNYERYSEEBEQHEEERPTNYSIKYNEEKRH 1172
 DB 1003 -----GVRPLVPGTGS----- 1013

QY 1173 VDQPIDYSLKATDIPSSQKQSFSSKSSQSGSKTEHSSSSSTSTPSSNAKRONQLH 1232
 DB 1014 -----PGARKQAW----- 1021

QY 1233 PSSAQRSGRQPKAATCKVSSINQETIQTVCYEDTPICFSRCSLSSLSAEDBIGCNQT 1292
 DB 1022 -IPADLSKVEKILVAPSL-PIASKVQKLVQAGPNSLRCSLSSLSSTGHAVPQAE 1079

QY 1293 TOEADSNANTLOIAIEKIGITRSAB-----DPVSEVPASVQHPTKSRLOGS 1340
 DB 1080 NLDSDS-----LEGLEAGPGEALGEAMRASGSTSLPVS-IPA-----PQRGRSR----- 1125

QY 1341 SLSSSARHKAVERFSSGAKSPKSGAQTQPKPPEHYVOETPLMFSPCTSVSSLSDFERS 1400
DB 1126 GLGVEDA-----TFSSSENVCQETPLVLSRCSVSSLSGFSERS 1165
QY 1401 TASSVQSPBCSMGWSIIPSLDPSQGTWPPSRKTPPPPPQTAQTKR-----EVPK 1454
DB 1166 TASSIFSPBCSLGSGTVPBELPSPQGTWPPSRKTPPPPPQPETQSLOWESYVK 1225
QY 1455 NKAPTAEKREGSPKQAAVAVQVQVLPDADTLHFATETPTDGFCSLSLSALSLDEP 1514
DB 1226 RFLDIADCRERQCPSELSDAGSVR-----FTVEKPDENFSCASSLSALHEL 1273
QY 1515 FIOKDVLRIMPPV--QENDNGNETESEQKSNENQKEAKETIDSEK---DILLDDSDDD 1570
DB 1274 YVQDVEURLRPPACPERAVG-----GCHRRRDEAASRLDGPAPAGRSARATDK 1324
QY 1571 DIEILEECIIISAMPTKSRKAKPAQATASKLPVPVARKPSOLPVYKLPSONRLQPOKHV 1630
DB 1325 ELEALRECLGAAMPARLAK-----VASALVP--GRSLPVPVYMLVPAPAR----- 1368
QY 1631 SFTPGDDMPVYCVGEGTINFINSTATSLDILIESPPNELAAGEVURGGAQSGEFKEDTI 1690
DB 1369 ---GDD--SGTDSAGTVPNFSSAASLSDTLQGPSRDKPAGPGR-----OK----- 1411
QY 1691 PTEGRSTDEAOGKTSVTIPELQDNKAEEDGIIAECINSAAMPKGSKHKPRVKIMDOV 1750
DB 1412 PT-GGAAPARQ-----TRSRP----- 1427
QY 1751 QOASASSAPNQLDQKKKPTSPVKPI PONTY-RTVRKNADSKNNLNAERFSDNK 1809
DB 1428 KAAGAGKS-----TEHTRGFCNRRAGLEPLSRPOSARSNR 1463
QY 1810 DSKONLKNNSKDFNDKLPNNEDRVGSPAFDPSPHYTPIEGTVCFSRNDSSLSLDFDD 1869
DB 1464 DSSCOT-----RTRGGLQSLCLTTPTEAVYCF-----YDS 1496
QY 1870 DDVILS-----REKAEIRKAKENKESBAKYTSHTELTSNQQSANKTQAIKQPIN---R 1920
DB 1497 DEEPATAPPERRASAIPLALKREKPAKCK-----ETPSRAAQPATILPVR 1541
QY 1921 GQPKILOKQSTFPQSSKDI PDRGATDEKLNFAIENTPVCFHNSLSLSLSDIDQNN 1980
DB 1542 AQPLI-----VDETPCYSLTSSASLSLSE----- 1566
QY 1981 NKNEPIKETBPDDQSGFBKPOAGYAPKGFHVHEDTPVCFSRNSSLSDIDSEDLLO 2040
DB 1567 -----PEA---PEQPANHARGEQGSKD-----SSPSPRAEEELLO 1600
QY 2041 ECISAMP-----KKKPSRLKGNKHSRPNMGGILGEDTLDLKDIOQPDSE-H 2090
DB 1601 RCISLAMPRRRTQVFGSRRRKPRALRS-----IRPTEIT-----QRCQBEVA 1643
QY 2091 GLSP--DSENFQKAIQEGANSIVSLSHQAACLSRQASSDSLSLSKSGISLGSPPF 2148
DB 1644 GSDPASDIDSVEWQAIQEGANSIVTLHQAAKASL--EASESDLSLSLVGVSAGSTL 1701
QY 2149 HLTPDQEBKFTSNKGRPIILKPEKSTLETKEI-ESKSGIKGKGVYKSLITQKVR-SN 2206
DB 1702 Q-----PSKLRKGRKPAABAGAWRPEKRGTTSTK-----INGSPLPLN 1740
QY 2207 SEISQMKOPLQANPSISRGTMTHIIGVNSSSSTSPVSKGPPPLKTPASKSPSEGT 2266
DB 1741 GPEKARGQJOMAGESTMLRGKTVIY-----SAGPASRTQSKG--ISGPTTTPKKTGTS 1792
QY 2267 ATTSPRGAKPSKSELSFVARQTSIGGSKAPRSRSGRSDTTPRPAQPLSRPIQSPGR 2326
DB 1793 GTTQETV-----TRAPSEQQRSRLHRPGKISELALALRHPPR 1831
QY 2327 NSISPRNGIIPPNKLSQIPRTSSP-----STATSKSG-----SKMYSYTFGQMGOQON 2377
DB 1832 SATPPARLAKTPSSSSQTSFASQPLPRRSPLATPTGGFLPCPGGSLVPKSPARAL----- 1897
QY 2378 LTKQTGLSKNASSI-----PRSESASKGLNQMNNGNGA-NKKVELSRMS 2420

RESULT 4

O95996 PRELIMINARY; PRT; 2303 AA.
ID O95996 AC O95996; DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE APCL protein.
GN APCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP Koyama K., Nakagawa H., Nakamura Y.;
RA "APCL exon14."; NCBI_TaxID=9606;
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Koyama K., Nakagawa H., Nakamura Y.;
RT "APCL exon14."; NCBI_TaxID=9606;
RN [2]
RP SEQUENCE FROM N.A.
RA Koyama K., Nakagawa H., Nakamura Y.;
RT "APCL exon14."; NCBI_TaxID=9606;
RN [2]
RP SEQUENCE FROM N.A.
RA Koyama K., Nakamura Y.;
RT "Identification of a brain-specific APC homologue, APCL, and its
interaction with beta-catenin.";
RL Cancer Res. 58:5176-5181(1998).
DR EMBL; AB022529; BAA75469.1; JOINED.
DR EMBL; AB022518; BAA75469.1; JOINED.
DR EMBL; AB022519; BAA75469.1; JOINED.
DR EMBL; AB022520; BAA75469.1; JOINED.
DR EMBL; AB022521; BAA75469.1; JOINED.
DR EMBL; AB022522; BAA75469.1; JOINED.
DR EMBL; AB022523; BAA75469.1; JOINED.
DR EMBL; AB022524; BAA75469.1; JOINED.
DR EMBL; AB022525; BAA75469.1; JOINED.
DR EMBL; AB022526; BAA75469.1; JOINED.
DR EMBL; AB022527; BAA75469.1; JOINED.
DR EMBL; AB022528; BAA75469.1; JOINED.
DR EMBL; AB012162; BAA34611.1; JOINED.
DR HSPP; Q02248; 3BCT.

[illegible]

QY 1931 STFPQSKDIPRGAATDEKLO-NFAIENTVCFSHNSLSLSLSDIDQENNNKENPIKE 1989
D 1546 KEAPAPSAAP--AAPPARTQPSLIADETPCYSLSSASSLS-----EPPE-- 1591
QY 1990 TEPPDSQCEPKQASGVAPKSFHVEDTPVCFSRNSLSLSLSDIEDDLLOECISSAMPK 2049
D 1592 SEPPAVHPRGEPAVT-----KDPGGGRRDSSSP-----RAAEELLORCISALPR 1639
QY 2050 KKKP--SRLLKGNKHSRPMGGILGEDITLKDIOIRP-----DSEHG-----LSPDSEN 2098
D 1640 RRPVPSGLR---RKKPR-----ATRLD-----ERPAEGSRGRGEBEAGSDRASDLDS 1683
QY 2099 PDKAIQAGANSIVSLHQAACCLSQASDSLSLSKSGISLGS-----PFLHPDQ 2154
D 1684 VEWRAIQAGANSIVTLWHAACAA--TREASESDSLSLFSVLSVSGSTLQPPPKRGRQ 1740
QY 2155 BEKPTSNKGRILPKPGKSTLETKKIBESGKIGKGGKVKYSLITGVRGNSEISGQMK 2214
D 1741 AEGEMGSARP-----EKGAASVKTSGSRSPAGPEK-----PRGT 1777
QY 2215 QPLQANMPSISGRMTIHI-----PGVNSSSSTSPVEKGPAPKTPA--SKS 2260
D 1778 OKTTGPAVLGRVTVIYVPPAPRAQPKGTGPRATPRKVP-----PCLAQAPAAK 1832
QY 2261 PSEGOTATTS--PRGAKPSVKSELSPVARTOISGSSKAPSRSGSRDSTPRPAQPLSR 2319
D 1833 PSPGQORSRSLHRPAKTSBELATLSQPPRSATPPARLAKTPSSSSQTS-----PASQPLPR 1888
QY 2320 ---PIQSPERNISPRNGIIPPNKLSQLPRTSPSTASTKSSGSKMSYTSPOKMSQQ 2376
D 1889 KRPPVTQAA-----GALPGGASPVKTPARTLLAKQHK-----TQSPVRIPPMQ 1934
QY 2377 NLTKOTGLSKNASSIPRESASKGLNOMNNGA--KKVLSRMSSTKSSGSESERDRERP 2435
D 1935 RPAPR--GPPPLARVP--EPGPRGAGTAGGARGGLVLRVASALLSSGES--SDRS 1989
QY 2436 VLVROSTIKAPSPTLARKLEESAFESLSPSRPASPTRSQATPVLSPLPMSLS 2495
D 1990 GPRRLTIKE--SPLARRSELSSAESASAPOGASPRGR-----PALPAVFLCS 2040
QY 2496 HSSVO-----AGWKLPNLSPTEYNDGRPAKHDTARSHSPSLPINSQTM 2547
D 2041 SRCEELRAPRQGPAPARQPPAAPS--PGERPAR-----RTTSPPSLPV--RAPAA 2091
QY 2548 KREHSHKSSSLPRVETWRTGSSSILSASSSESEKAKSEDEKHVNSISGKQSKENQVS 2607
D 2092 RPETVKRYASLPHISVARPPDGAAPASADAARRSDGPRPL-----PRVA 2140
QY 2608 AKG--TWRIKENEK-----SPTNMTSQTSSGATNGAESKTLIYQMAPAVSXTEDVWVRI 2661
D 2141 APGTTWRRIRDEVDPHILRSTLPATLPLRGSTPEDAPA-----GPPPKTSDAVVQT 2193
QY 2662 EDCPINPRSGESPTGNT--PPVIDSVSEKANPNIKDSKDNQAKQNVGNGSVPMRTVYLE 2719
D 2194 EVAAPKTNSTSPSLETREPPGAPAGQLSLGSDVDGSLAKAPI---SAPFVHEGLG 2250
QY 2720 NRLNFIQVADPDQGTBIKPCQNNPVPVSETNESSIVERTPFSSSSSKSSPSGTVAA 2779
D 2251 VAVGGF-----PASRGSPSR--SA 2268
QY 2780 RVTPEYNAPSPRKSADSTSA 2800
D 2269 RVPEFNYPSPMVVAATDSDA 2289

RESULT 5

Q8BRD8 PRELIMINARY; PRT; 489 AA.
AC Q8BRD8;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)

DE Adenomatosis polyposis coli.
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; PubMed=12466851;
RX MEDLINE=22354683; The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK045053; BAC32198.1; -.
DR MGD; MGI:88039; Apc.
DR GO; GO:0005737; Cytoplasm; IDA.
DR GO; GO:0005634; Cnucleus; IDA.
DR GO; GO:0008013; F-beta-catenin binding; IDA.
DR GO; GO:0009522; Anterior/posterior pattern formation; IMP.
DR GO; GO:0009798; P-axis specification; IMP.
DR GO; GO:0009533; P-dorsal/ventral pattern formation; IMP.
DR GO; GO:0016055; P-wnt receptor signaling pathway; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR SMART; SM00185; ARM; 1.
SQ SEQUENCE 489 AA; 5521 MW; F7C9949868B01993 CRC64;

Query Match 15.3%; Score 2232; DB 11; Length 489;
Best Local Similarity 93.0%; Pred. No. 6.3e-102;
Matches 441; Conservative 11; Mismatches 20; Indels 2; Gaps 2;
QY 1 MAASVDLLKQVEALKMENSRLQLEONSHTLKTLEASNMKEVLKQSGSIEDEAM 60
D 1 MAASVDLLKQVEALKMENSRLQLEONSHTLKTLEASNMKEVLKQSGSIEDETM 60
QY 61 ASSGQIDLLERLKLNLDSNFPFVKLRKMSLRSYSGSREGSVSSRGSCSPVPMGSFPR 120
D 61 -TSGQIDLLERLKEFNLD-SNFPFVKLRKMSLRSYSGSREGSVSSRGSCSPVPMGSFPR 118
QY 121 RGFVNSRESTGYLEBELEKERSLLADLKBEKEKWYTAQLQNLTKRIDSLETFENFSL 180
D 119 RTFVNSRESTGYLEBELEKERSLLADLKBEKEKWYTAQLQNLTKRIDSLETFENFSL 178
QY 181 QTMTRQLEVEARQIRVAMEEQLGTCDQMEKQAQRRIARIQIIEKDIILRIQLLOSOAT 240
D 179 QTMTRQLEVEARQIRVAMEEQLGTCDQMEKQAQRRIARIQIIEKDIILRVQLLSQAA 238
QY 241 EAERSSQNKHETGSHDAERQNEGGYGEINMATSNGQSGSTTRMDHETASVLSSSSTHSA 300
D 239 EAERSSQNRHDAASHEAGRQHEGHEGVAESNTAASSSGQSPATRVDHETASVLSSSSTHSA 298
QY 301 PRRLTSHLGTKEWYSLLSMLCTHDKDDMSRTLLAMSSSDSCISNRQSGCLPLLIQLL 360
D 299 PRRLTSHLGTKEWYSLLSMLCTHDKDDMSRTLLAMSSSDSCISNRQSGCLPLLIQLL 358
QY 361 HGNDKDSVLLGNRSGKEARASAAALHNIHSGPPDKRGRREIRVLHLEQIRAYCETC 420
D 359 HGNDKDSVLLGNRSGKEARASAAALHNIHSGPPDKRGRREIRVLHLEQIRAYCETC 418
QY 421 NEWQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGLQA 474
D 419 NEWQEAHEQGMDDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGKIK 472

RESULT 6

Q9UBZ1 PRELIMINARY; PRT; 733 AA.
AC Q9UBZ1;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)

Db 168 ----- 167
QY 359 LLHGNDKDSVLLGNSRSGKEARASAAALHNIHSQPDQKRRRIRVHLLEQTRAYCE 418
Db 168 ----- 167
QY 419 TCWEQOEAEHPGMDQKNPMPAPVEHICPAVCVLMKLSFDBEHRHANNELGGLQAI AEL 478
Db 168 ----- EL 169
QY 479 LOVDCMYGLTNDHYISITLRRVAGMALNLTGDEVANKATLCMKGCMRALVAQKSSSE 538
Db 170 LOVDYEMHKWTRDPLNLAL-RVAGMTLNTLTGDEVANKATLCARRGCM EAI VAQLASDE 228
QY 539 DLOQVLA SVLRNLWRADVNSKKTREVSGVKALMECALEVKKESTLKSVALNLSAH 598
Db 229 ELHQVSSILRLNWRADINKNLREAGSVTALVQCVRATKSTLKSVALNLSAH 288
QY 599 CTENKADICAVDGAALFVLTITYSQNTLAIIESGGILRNVSLLIATNEDHQILRE 658
Db 289 STENKAAICQVDGALGLFVLTITYSQNTLAIIESGGILRNVSLLIATNEDHQILRE 348
QY 659 NNCLOTLLOHLKSHSITVSNACGTLNLSAENPKDOEALMDGAVSKMLIHSKHOMI 718
Db 349 HNCLOTLLOHLKSHSITVSNACGTLNLSAENPKDOEALMDGAVSKMLIHSKHOMI 408
QY 719 AMGSAALNLMANRPAYK-DANIMSPGSSLPVSLHVRKQKALEAEADQLSETFDNID 777
Db 409 AMGSAALNLMANRPAYK-DANIMSPGSSLPVSLHVRKQKALEAEADQLSETFDNID 468
QY 778 NLSPKASHRSKO-----RHKOSLYGDIVFDNTHDDNRSNF-----NTGNMTVLS 823
Db 469 KOGPPAAEAATKPLPLRLHLGLAQDYASDSCFDDDDAPSSLA AAAAATGEPASPAALS 528
QY 824 PYLNTLTVLPSSSSSR---GSLDSSSEKDRSLERERIGLGNHYHPATENPGTSSKRGLOI 880
Db 529 LFLGSPFFLQCGAQAATPTRRCGKEAKDTSG-----AAVAKAKAKL 572
QY 881 STTAQIAKMEVEIAHTSQDRSGSTTELHLCTVDERNALRSSAAHTHS---NTNYFT 938
Db 573 ALAVARIDQLVEDISALHTSDSDSFLSS-----GDPQZAPREGRAQSCSPCKGPEGG 626
QY 939 KSESNRNTCPYAKLEYKRSNDLSNVSSSDGYGKGQMKPSTIESVDESKFCSYG 998
Db 627 RREAGRAHPLRLKAAASJNDSLSGASDGYCPREH----- 667
QY 999 QYPADLAHKIHSANHMDDNGELDTPINYSKYLSDEQLNSGRQSPSQNERWARKHIIE 1058
Db 668 LPCELA-----ALASREDP----- 682
QY 1059 EIKQSEQRQSNQSTIYPVYTESTDDKHLKQPHFGQOECVSPYRSGANGSETRVGSN 1118
Db 683 -----RCGQPRPSR----- 691
QY 1119 HGINQWSQLCQEDDYDDDKPTNYSERYSEEEHEEERTNYSIKYNEEKRHVDQPID 1178
Db 692 -----LOLDPFGCAEPAREATSADARVT---IKLSPYQHV----- 727
QY 1179 YSLKATYDIPSSQKQSFSSKSSQSKTEHMSSESTTTPSNAKRONQLHPSSAQS 1238
Db 728 -----PLLEGASRACAEPLAGFGI-----SPGARKQAWL---PADH 760
QY 1239 RSGQPKAATCKVSSINQETITQYCVETDTPICFRCSLSLSAEDIEGCGNQTQADS 1298
Db 761 LSKVPEKLAAPL-SVASKALQKLAQEGGULSRCSLSLSA-GRPGESEGGDLDS 818
QY 1299 ANTLQIAE-----IKEKIGTRSAEDPVSEVPVAVSOHPRTKSRLOQSSLSSEARHKAVEF 1354
Db 819 DSSLEGEAEAGFIEALDSTWRAPGATSLPVAIPAPR-----RNRRGLGVEDA----- 867
QY 1355 SSGAKSPKSGAQTPEPPEHYVQETPLMFRCTSVSSLDSEFERSIASSVQSEPCSGMV 1414
Db 868 -----TPSSSESNYVQETPLVLRCSVSSLSGSEFSPSTASSIPSEPCSGHG 914

QY 1415 SGIITSPDLSPDSCQTMPPSRSKTTP--PPQTAQTKREVVPKNKAPTAERESGPKQAAV 1472
Db 915 SGIITSPDLSPDSCQTMPPSRSKTTPPLAPAPQ-----PPEATQPSLOWESVYKFLD 967
QY 1473 NAAVQRVQVLP---DADTLHTEATEPDCGSSSLSALSJLDEPFTOKDVELRIMPVQ 1529
Db 968 IADCRERCLPSELDAJS-VRFVTEKPDENFSCASSLSALALHEHYVQDVEURLLPSAC 1026
QY 1530 ENDNGNTESEQPKESNENQEKAEKTIIDSEKOLLDDSDDDDIIELEECIISAMPTKSSR 1589
Db 1027 PERGGAGGAGLHFAHRRRGGAGAGFSPSR-----RRGOELLEFLRECLGAAPARLRK 1081
QY 1590 KAKPACTASKLPPPVARKPSQLPVYKLLPSONRLQPKHVSFTPGDDMPRVTCVEGTPI 1649
Db 1082 -----VASGLVP--GRRALPVVYMLVPAPAPQ-----EDDSDCTDSAEGETPV 1122
QY 1650 NFSTATSLDTTIESPPNELAAGEVGRGGAQSGEFEKRDITPTEGRST-----DEAQQGKT 1705
Db 1123 NFSSAASLSDETLOGPRDQPGGAGR-----QRPTRPTSAQAMGHRHKAGAGR 1174
QY 1706 SSVTIPELDNKABEGDILAEICINSAMPKSHKPRVKKIMQVQOASASSAPNKNQL 1765
Db 1175 SABOSPGAGKXRA-----GLEPLG----- 1194
QY 1766 DGKKKPTSPVKPIPQNTYRTRVRKNADSKNNLNAERVSFDMKSKQKQKNNKDFND 1825
Db 1195 -----RPPSAA-----DKDSKPG----- 1209
QY 1826 KLPNNDRVGRSFAFOPPHYTFIEGTPYCFSRNDSL 1862
Db 1210 -----RTRGDGALQSLCTTPTTEAVYCYGNDVS 1239
RESULT 8
P91667 PRELIMINARY; PRT; 2416 AA.
ID P91667 AC P91667;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE ADENOMATOUS polyposis COLI.
APC OR D-APC OR CG1451.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=37144426; PubMed=8990193;
RA Hayashi S., Rubinfeld B., Souza B., Polakis P., Wieschaus E.,
RA Levine A.J.;
RT "A Drosophila homolog of the tumor suppressor gene adenomatous
polyposis coli down-regulates beta-catenin but its zygotic expression
is not essential for the regulation of Armadillo.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:242-247(1997).
DR EMBL; U77947; AAB41404.1; -;
DR PIR; T13825; T13825.
DR FlyBase; FBgn0015589; Apc.
DR GO; GO:0008013; F-beta-catenin binding; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg. 5.
DR SMART; SMO0185; ARM; 5.
DR PROSITE; PS0176; ARM_REPEAT. 2.
SQ SEQUENCE 2416 AA; 261405 MW; 4DE3A10BE10B42A9 CRC64;
Query Match 12.2%; Score 1780.5; DB 5; Length 2416;
Best Local Similarity 25.4%; Pred. No. 9.le-79;
Matches 720; Conservative 329; Mismatches 847; Indels 935; Gaps 100;
QY 243 ERSSQNHKTGSHDAERQN-EGQGVGEINMATSNGQSGTTR-----MDHETASVLSSS 295

Db 129 ELREMEHRLDRNFRQSAQQOQLDELPRNGGGSPASAGRPSPRSKEPSYTLRSFLDGD 188
 QY 296 STHSAPR-----RLTSH-----LGTQVEMVYSLLSMLGTHDKDMSRTLLAM 337
 Db 189 APAPAPRLPKGAANTTSFERYTSSAVEATLGSKEVCYVSLLSMLGNSDPLEWAKFLEL 248
 QY 338 SSSQDCSICMRQSGCLPLIQLIHLGHNDKDSVLLGNSRSGKEARASAAALHNIHISQPD 397
 Db 249 SGAQSCATLRRSGCMFLVQMHAPDND-----QEVKCKAEQALHNVHSHPDE 298
 QY 398 KRGRREIRVLHLEBOIRAYCETCWHEAHEPGM--DQDKNMPAPVEHQICPACVLMK 455
 Db 299 KAGRREAKVRLLDQIVDYCSFKTLTLOSQGEAIAODSDRHPL-----AALSSLMK 349
 QY 456 LSFDEEHRHAMNELGLLOAELQLVDCEMYG-LTNDHYSITLRRYAGMALTNLTFGDVA 514
 Db 350 VSDEEHRHAMCELGAHALPVLHDHVAHVGPKPEDQCCNSLRRYALMALTNLTFGDEN 409
 QY 515 NKATLCMKCMGCMRALVAQLKSESDILOQVITASVLRNLSWRADVNSKTLREVSUKALME 574
 Db 410 NKALLCGQKFMALVAQLDSAPDQLQVTASVLRNLSWRADNMMKAVLNEIGTIVTALAL 469
 QY 575 CALEVKKESTLKVLSALWNLASHCTENKADICAVDGAFLAVGLTLYRSQNTWLAIES 634
 Db 470 AAMNRSNTLKAISALWNLASHCTENKAEFCVAVDGAFLAVGLMSYEGPSKTLKIEN 529
 QY 635 GGGILRVSSLIATNEPHROILRENNCLOTLLQHLKSHSLTIVSNACGTLWNLARSNPKD 694
 Db 530 AGGILRVSSHIAVCEPYQILRCHNCALILLOQLKSESLTVVNSCGTLWNLARSARD 589
 QY 695 QEALWDMGAVMLKNLHSHKMTAMGSAALANLMAANRPAYK----DANIMSPG-SSL 749
 Db 590 QKFLWDMGAVMLKSLHSHKMTAMGSAALANLMAANRPAYK----DANIMSPG-SSL 749
 QY 750 PSLVRKQKALEABDLAQHSETFDNIDNLSPKASHRSKORHKQSIYGDYVFDNRHDDN 809
 Db 650 PTLKARAKALQELGEBRHTAETCDNLD-----KLDKERASSSSRRHP----- 677
 QY 810 RSDNFTGNMTVLSPLYNTTVLPSSSSSRGLDSSRSEKDRSLERERGICLGNYPHATEN 869
 Db 678 -----TGG-----KLDKERASSSSRRHP----- 695
 QY 870 PGTSSKRLQITTAQIAKVMEEVSAIHTSQEDRSSTGTELHCVTDERNALRRSSAAH 929
 Db 696 -----APLIR-----SAMLTKSESROSVYSAKSDCAVDH----- 725
 QY 930 THSNYFTKSENENRTCSMPYAKLEYKRSNDSNLSSVSSDDGYGKGQWKPSIESYSED 989
 Db 726 -----LIRSASASDAHRK-----VKPKITDFDLE 749
 QY 990 DESKFCYGVYPADLAHAKHISANHMDNDGELDTPIYNSLYKDEQLNSGRQSPQNERW 1049
 Db 750 ME-----QTEATEGQPIDYSVKYS----- 769
 QY 1050 ARPKHIIIEIKOSEQRQSNQSTPYVYTESTDDKHLTKQPHFGQOECVSPYRSRGANG 1109
 Db 770 -----ENAKTKSTY----- 778
 QY 1110 SETNRVSGHNGINQVQSLOQEDDYEDDKPTNYSEYSEBQHEBERPTNYSIKYNEE 1169
 Db 779 -----QETDL--DQPTDFSLRYAE--NOTESLDDISGAGGQ 811
 QY 1170 KRHYDQPIDYSLKYATDIP--SSQXQSFPKSSGQSKTEH-----MSSSENSTTPSS 1223
 Db 812 KSTITPP-----AETVPEKSEQELILLDDSVKCYQTEDETPYVLSNAASVTDLRVA 864
 QY 1224 NAKQONQLHPS--SAQSRSGOPO--KAATKYSSINQEIOTYCYVEDTPICFSCSSLS 1278
 Db 865 KADREAEVKPEVRVTSKEGAPKPLKLSQCGSGSYTPEKPIVCEBGTFCYFSRVDLS 924
 QY 1279 SLSSAEBIG-CNQTTQBADSANLQIAEIKETGRSAEDPVSEVPASVQHPRTKSSRL 1337

Db 925 SL-----DESKANQAIVGTD-----ADIKPKLEKQEQESQPAEQVLTKPTQANSAL 973
 QY 1338 QGSSLSSESARHKAVERFSSGAKSPKSGAOTPKSPPPHYVQETPLMPSPRSTSVSSLDSE 1397
 Db 974 -----ETPLMPFSRRSSMDSLVHDP 992
 QY 1398 SRSTA-----SSVQSEFSCGMVSGIISPSDLPSPGOTMPPSRSKTTPPPQTAQTKREV 1452
 Db 993 DVTVANDCDSSVSD--FSLASGVISFSEIPOSTQSMFQS----- 1033
 QY 1453 PKNKAPTAEKRESGPKQAANVAQVRVLPDADTLHLHFAETSTPDGFCSSLSALSULD 1512
 Db 1034 PRNSVAGSGQVDSPPVWIPASLQPLRSVFE--DDLSSFNVEHTPAQFSTATSLNLSI-- 1091
 QY 1513 EPFIQKQVELRIMPVQENDNGNE-----TESEQPKESNENCKE-AEKTID 1558
 Db 1092 -----VDDEKAPAVWTEDEDELLANCINMGOMKQKTEAVKSTVNVSEVDVAEETIR 1144
 QY 1559 S-----EKD-----LLDD--SDDDDI----- 1572
 Db 1145 SYCTEDTPALLSKVPSNTNLSVISWSTDPKDATAGQAQVMAHQLSDDVSSNASDCGAS 1204
 QY 1573 -EILEECIISAMPYKSSRKAAPKPAQTAASKLPPVARK-PSQLPVYKLLPSQNRLOPKHV 1630
 Db 1205 GHLQOCTRDGM-----KKPLGEATSDPIAMLRGGNELPGY--LPS----- 1244
 QY 1631 SPTPGDDMPRVYCVGEGTPIINFSTATSLDLTIES-----PNNELAAAGRGVGGAGSEF 1686
 Db 1245 -----ADEMVK-FLVEDSPCFVSVGSLNLTGSSLVGPAVLKETE- PSSADQNPEMKR 1298
 QY 1687 RDTIETGRSTDEAOGGKTSSVTIPELDDNKAESGDILAE----- 1727
 Db 1299 SLANRSKRRPPHQDDSLSLSDSDSDTLLSQAIAAGCNRPKSLGFSNNGKRSSL 1358
 QY 1728 -----INSAMPKGSHKPRVKIMDOVQQAASASSAPNKQ-----LDG--KK 1769
 Db 1359 SSSOPIAINAATASSLNSAMTVRSQOQESYSSVSDSDNDNSKSLFELCILKGMKYT 1418
 QY 1770 KKP-----TSPVKPIQNTYRT-----RVKNADSKNLAER 1803
 Db 1419 KEPGARAQMOEQPIVSGSSVQSNPSLKQDPSLPVQLPSSQGVKQRHHHHHHHRRER 1478
 QY 1804 VPSNDKSK-KQNLKN--SKDFNDKLPNNEDRVGSGFAPDSPHYTPLEGTPYCFERN 1860
 Db 1479 ---EKDEKLQECINTGISKKIN-AVPKNV--LATSAALEPCH--PMAAT-----T 1523
 QY 1861 SLSSLDFODDVDLSREKALRKAKENKESAKVTSTHTLTSTNOQSANKTQA-TAKOPI- 1918
 Db 1524 SASALSTAAPDVE---QKAH--ATSNPQOQSTHPSHILPNPIDATVTDARGPAA 1577
 QY 1919 -NRGQPKPILOKQSTFPOSSKIDPRGAATKELQNFAIEN-----TPVCF----- 1963
 Db 1578 PNQNGNASQNGLET-ATGSKDLSDSESDSNQSFIMETMTVRLDSALNETCISGASEK 1636
 QY 1964 -----SHNSLSLSLSDIDQENNNKENEPIKETEPDPS 1995
 Db 1637 HKDPDLMLKSVERTLMEVFTAELRSSHSHSS-----NSHKNNSSNTWNESTCPND 1691
 QY 1996 QGEPSPQASGVAPKSFHVEDTPVCFNRNSLSLSLD-----SEDDLOECISSAMPKKK 2052
 Db 1692 VSFPSVSQTA-----PV-----LASLSDATEDARSHELLEITPTNEQQ 1732
 QY 2053 PSRLKGDNEK---HSPRNGGILGEDITLDLKDQRPDSEHGLSPDSENFDMKAIQEGA 2108
 Db 1733 PESLEGETDTLVNGHADS--SGSSGGLNFQ---GGQVQAGVRLEPQELLENGTSASIMT 1789
 QY 2109 NSIVSSHLQAAAAACLSRQSSDSD-----SILSL-----KSGI-SLGSPP--HUTP 2152
 Db 1790 NSTMIAFEARALAEALLQPAATDDTTMTFSLNSLDLNDIRP7SGMBSLNSCYQDSQP 1849
 QY 2153 DQEEKPF7SNKGPRLIKRGEKSTLETKIESEKIGKGGKVKYSLITGKVRNSENISGQ 2212
 Db 1850 SSLRQAMPS-KSPRFAKNFFANLVARALGHLAG-----SAESVNSSCNLDN 1897

530 AGGILNVSHIAVCEPYRQILRHQNCIAILLQOLKBSLTVVNSCGTLWNLSARSAED 589
 695 QBALWDMGAVMLKJLIHSHKHMIAMGSAALNLMANRPAYK-----DANIMSG-SSL 749
 590 QKFLWDMGAVPMLRSLIHSKHAMISEGSSSALNLIINFRPAQVNHQDPIARSMGLKAL 649
 750 PSLHVRKOKALEAEHDAOLHSTFONIDNLSPKASHRSQRKHQSLYGDYVDFVDTNRHDDN 809
 650 PTLZAKAKALQELGERHTAETCNLD----- 677
 810 RSDNFNTGNTVLSPLYNTTVLPSSSSRGSLDSSRSKDRSLERBERGIGLGNYPATEN 869
 678 -----TGG-----KLDKERASSSSRRHP----- 695
 870 PGTSSKREGLOISTTAACIAKWEVEIAHTSOEDRSSGTTTELHCVTDERNALRSSAAH 929
 696 -----APRLTR-----SAMLTKESRDSVYSAKSDCAYD- 725
 930 THSNTYNTKSENSNRCTCMPYAKLEYKSSNDLSNVSSSDGYGHRGOMKPSIESYSED 989
 726 -----LIRSASADAHRK-----VKPKITDFOLE 749
 990 DESKPCSYGQYPADLAHIAHNMDDNDELDTPIVNSLYKYSDEQLNSGRQSPQNERW 1049
 750 ME-----ODTEATEQPIDISVKYS----- 769
 1050 ARPKHIIIEIKQEQORSNQSTTVPVYTESTDDKHLKFOHPGQOECVSPYRSGANG 1109
 770 -----ENATKSTV----- 778
 1110 SETNRVGSNGINQVNSQLCOEDDEDKPTNYSERYSEEBEERPTNYSIKYNEE 1169
 779 -----QETDL-----DQPTDFSLRYAE-----NOIESDLOISGAPGQ 811
 1170 KRVHDOPIDYSLKYATDIP--SSQKQSFSSKSSGQSSKTEH-----MSSSENTSTPS 1223
 812 KSIITPP-----AETVPEKSEGEILLILDSDVKCYCTEDTPYVISNAASVTLDRVA 864
 1224 NAKRONQLHPS--SAQSRSGOPO--KAATCKVSSINQETIOTYQVEDTPICFSRCSLS 1278
 865 KADAEABVKEPREVTSKEGAPKLPKLSQCGSGSYTPEKPINYCEGTPGVFSRYDLS 924
 1279 SLSSAEDEIG-CNOTQEQASANTLQABIKETGRSAEDPVSEVPVAVSQHPRTKSSRL 1337
 925 SL-----DESKANQAIUGTD-----ADIKPLEKQEESOPAEQVLTKPTQANSAL 973
 1338 QGSSLSSESARHKAVFSSGAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVSSLDSP 1397
 974 -----ETPLMFSRRSSMDSLVHDP 992
 1398 SRSTA-----SSVQSEPCSGMVSGIISPSDLPDSPGOTMPPSRKTPPPPPQTAQKREV 1452
 993 DVDVANDCKSSVSD--FSLRAGSVISPEISDPTQSMPOQ----- 1033
 1453 PKNKAPTAEKRESGPKQAANAAVQRVQLPDADTLHFEATESTPGFCSSLSALS- 1511
 1034 PRNSVAGSGQVNDSPVWIPASIQPLRSVFE--DLLSSFVNEHTPAQFSTATSLNLSIV 1092
 1512 -DEPFIQXVELRIMPVQENDNGNE-----TESQPKESNENQKE-AEXT 1556
 1093 DDE-----KAPASVAEEDNEDELLLANCINMGORMKPTAEVKSVTNVNSEVDVAET 1143
 1557 IDS-----EXD-----LLDD--SDDDDI-- 1572
 1144 IRSYCTEDTPALLSKVPSNTNLSVSMSTDPKDATAGQAMVAHQSLDSDVSSNASDCGG 1203
 1573 ---BILBECHISAMPTKSSRKAQKPAQATASKLPPPVARK--PSQLPVYKLLPSQNLQPOK 1628
 1204 ASGHLLQOCIRDMG-----KKPLGEATSDPTAMLRGNGELPGY--LPS----- 1245
 1629 HVSTPGDDMPVYCVSGTPINRSTATSLSLDTIES-----PPNELAAGEGVGGAQSGEF 1684
 1246 -----ADEMMK--FLVEDSPCNFVSGLSNLTIVGSSLVGPAVOLKETE--PSSADQNPEM 1297

QY 1685 EKRDITPTEGRSTDEAQQGGKTSSVTIPELDDNKAEEGDILAE- 1727
 Db 1298 KAREGKEQVRRPHWQDDSLSLSDSEDDTNLLQAAGCNRPKSNLGFSSNGKRSS 1357
 QY 1728 -----INSAMPKGSKHPFRVKIMDOVOQOASAGSAPKNO-----LDG-- 1767
 Db 1358 SLSSSQPIAINAATSASSLNSAMTVRKSOQOESYSSVSDSDNDNOSKSLFELCILKGMV 1417
 QY 1768 KKKKP-----TSPVKPIQNTYEY- 1801
 Db 1418 KTKEPGARAQMOEQPIVSGSSVQSNFSLKQPDLPVQLPSSGOVKQRHHHHHHHRER 1477
 QY 1802 ERFVSDNKDK-KONLQW--SKDFNDKLPNEDRVGSPAFSPHYPHYPIEGTFCFSR 1858
 Db 1478 ER-----ERKDEKLLQECINTGISKIN-AVPKV--LATSAAALEPCH--PMAAT- 1523
 QY 1859 NDSLSLDFDDDDVDLSREKAEKAKENKESAKVTSTELTNSQOSANKTOAIKQPI 1918
 Db 1524 -TSASALSTAAPV-----EOKAHATSNPQ--QOSSTHPSHILPNPI 1563
 QY 1919 -----NRQPKPILOKQSTFPQSSKDIIPDRGAATDEKLOFAIEN----- 1958
 Db 1564 DAIAVTDTVRSAPAPNQGNGNASQNGLET--ATGSKDLDESDSDSNQSFIMETWRL 1622
 QY 1959 -----TPYCF-----SHNSLSLSLSDIQENN 1980
 Db 1623 DSALNETICISASEKHKDDPDLMLKSVERTWFTVSAEQLRSSSHHSS--NSHKN 1677
 QY 1981 NKENEPIKETEPDPSQGEPSKQASGYAPKS FHVEDTPVCFSRNSSLSLSID--SEDD 2037
 Db 1678 NSSNNTWNESTCPNDVSPSVQTA-----PV-----LASLSLDEDATEAR 1718
 QY 2038 LIQECISAMPKPKKPSRLKGDNEK-----HSPRNMGGILGEDLTLDLKDQRPDSEHGLS 2093
 Db 1719 SLHELIEITPNEQPSLEGETDVLNGHADSVSGSSGGLNFQ-----GGQVQVAGVLE 1775
 QY 2094 PDSENFOWKATQEGANSIVSLHQAACAAACLSRQASSDSD-----SLSL-----KS 2140
 Db 1776 PQLLFGTSSASIMTNTMTIAFEARALAEALLOPAATDDTTTMTFSLNLDLNRIPPS 1835
 QY 2141 GI-SLGGPF--HLTPDQEEKFTSNKGPRLKPEKSTLETETKLESKGIKGGKVKYS 2197
 Db 1836 GMEISLNCYQDHSQFSSLRQAMPS--KSPFRFARKVFPANLVARRALHLAG----- 1884
 QY 2198 LITGVRSNLSIGOMKOP-----LOANMPSISRGRMTIHIPGVNRSSSSTSPVSKGPP 2252
 Db 1895 -SAESVNSSCNLLDNIRKPPSLMDELDSMISVDLSIQSEVADGEODCSMAITISVSNYETA 1943
 QY 2253 L-----KTPAKSPSGEQATTS PRGAKPSVKSEL----- 2282
 Db 1944 ACDDQMTVLQSCFDEDEDATMNDYSSAESTPKHGSTFSPNRRSLTPDKRRRLTKDFKT 2003
 QY 2283 -----SPVARQTSOI-----GGSKAPSRGSRDSTPSRPAQOPLSRPI----- 2321
 Db 2004 YTIATSCMEAPEANETUQIEIVAAVPAVTPSPRANGRRGSAERYKTOLIECPALIQ 2063
 QY 2322 -----QSP--GRNSISGR 2333
 Db 2064 PQDDCPSEQLSSITRAMMQOFTFITDINIGHSOETCTDHPEDAGESPECDSNETESC 2123
 QY 2334 NG-----ISPENKLSOL-----PRTSSPSTASTKSGSGKSVYTPGROWSQNLTKQT 2382
 Db 2124 DQQEPDQLPPPSIVDLRTSVVKTLEPAFAVLRGRKKPAVPSVYMSQORN----- 2178
 QY 2383 GLSKNASSIPRSEASKGLNQMNANGANKKVELSRMSSTKS-----SGSESDSRERPV- 2436
 Db 2179 -----SNNAPSKKKTLSPTIAKESLVPGGSGVRLPAKKKPTP 2216
 QY 2437 -----LVQSTFIKEAPSTLRLKLEESASFESLSPSSPASPTESQATVLSLPSL 2488
 Db 2217 PPEPAPARLERQGTVKD-----EPTNSNVQVPVVE--TK 2249

QY	2489	PDMVSLSTHSSVQAGGWRKLPDPNIPSTIEYNDGRPAKXHDHIAKSHSSPSRLPINRSGTWK	2544
Db	2250	PAQTSPTPHRA-----SKLPTKKGTA---SGGSFSK-----AGSPKRIPLAPA---R	2289
QY	2549	REHSKHSSLPYRSTWRRTQSSSSILSSASSESEKAKSEDEKHVNSISGTQSKENQVSA	2608
Db	2290	RMTFQRAETSLRLAAGKSHAASRVSGRVSTTFPSKSNLNGSSAAAAAKINHAQS	2349
QY	2609	K--GTRWKIKENEFPTNS---TSQTVSGATNGAASKTLIYOMAPAVSKTEVDVWVRIED	2663
Db	2350	RIANIWKRVDEAKTKOSSNLRTQKTKSSNMLNANGTKPTLLR-----SSTFD-----	2397
QY	2664	CPINPRSGRSGTGTNPVIDSVSEK	2669
Db	2398	-----NTPSTAGGVKSK	2409
RESULT 10			
Q9UEM8			
ID	Q9UEM8	PRELIMINARY;	PRT; 1685 AA.
AC	Q9UEM8;		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	APC2 protein (Fragment).		
GN	APC2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	KCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RF	MEDLINE=99147086; PubMed=10021369;		
RA	van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,		
RA	Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;		
RT	"Identification of APC2, a homologue of the adenomatous polyposis coli		
RT	tumour suppressor.";		
RL	Curr. Biol. 9:105-108 (1999).		
DR	EMBL; AUI31187; CAA10317.1; -.		
DR	GO; GO:0005578; C:extracellular matrix; IEA.		
DR	GO; GO:0004222; F:metalloendopeptidase activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000225; Armadillo.		
DR	InterPro; IPR001818; Pept_M10A_M12B.		
DR	Pfam; PF00514; Armadillo_seg. 2.		
DR	PROSITE; PS00546; CYSTEINE_SWITCH; 1.		
FT	NON TER		
SEQ	SEQUENCE	1685 AA; 175540 MW; B9B81923F3912F77 CRC64;	
Query Match			
Best Local Similarity 11.7%; Score 1705; DB 4; Length 1685;			
Matches 641; Conservative 263; Mismatches 664; Indels 682; Gaps 86;			
QY	654	QILRENNCLQTLQHLKSHLSLTVSNACGTLNLSARNPKDOEALWDGAVSLKNIHLS	713
Db	1	QVLRDHNCLQTLQHLTSLTVSNACGTLNLSARSARDQELWDGAVGNLRNLVHS	60
QY	714	KHKMTAMGSAALRNLMANRPAYK--DANIMSPGSLPSLHVKKQKALEAELDAQHSET	772
Db	61	KHKMTAMGSAALRNLLAHRPAKHQAAATAVSPGSCVPSLYVKRQKALEAELDASHLAQA	120
QY	773	FNIDNLSPKASHRSKQ-----RHQSLYGVYVPTNPHDNRSDN-----FNTG----	817
Db	121	LEHLEKQGPAAEAATKELPLRHLDDGLAQYASDGCFFDDDDAPSSAAAAATGEPAS	180
QY	818	----NNTVLSPYINTVLPSSSSRGLSDSRSEKDRSLERBERGICLGNYPATENPGTS	873
Db	181	PAALSFLGSPFLQQAARTPTPRG--GKEAKDTSGE-----AAVA	222
QY	874	KRGLCISTAAQIAKVMEEVAIHSTSQEDRSSGTTLHCVTDERNALRSSAAHTHS-	932
Db	223	AKAKAKLAVALARIQOLVEDISALHTSSDSDSFLSS-----GDFGQAPREGRAQSGSP	276

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yao S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Yu X., Bienz M.;
RT "A new Drosophila APC homolog concentrated in apical adhesion zones of
RT epithelial cells";
RL Nat. Cell Biol. 0:0-0 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
RT tumour suppressor";
RL Curr. Biol. 9:103-108 (1999).
DR EMBL; AF003746; AAF56249.1; -
DR EMBL; AF113913; AAD40227.1; -
DR EMBL; AF091430; AAD20985.2; -
DR FlyBase; FBgn0026598; Apc2.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 6.
DR PROSITE; PS01176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116702 MW; 20C8F5F6121888F8 CRC64;

Query Match 8.4%; Score 1222.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 8.6e-52;
Matches 429; Conservative 178; Mismatches 379; Indels 647; Gaps 50;

QY 329 DMSRTLLAMSSQSCISMROSGCLPLLIQLLHGNKDSVLLGNRSGSKEARARASAAH 388
DDB 12 ELTRNFELSINPECTALRSSDCIQLLVQILHANDE-----GLSTAKKYASQALH 62

QY 389 NIHSQPDKGRHREIVHLLLEQIRAYCETC-NEWQBAEP-GWDOQKNMPAPVEHQI 446
DDB 63 NIVHNPEKORORVKMLRLDQLDILYCNFLHTQLQSGGGAIAADDEDRHPL----- 114

QY 447 CPAVCVLMKLGDFDEHRRHAMELGLOAIABLLQVDCBMYG-LTNDHYSITLRYAGVAL 505
DDB 115 -AAMKLLMKASFDEHRTQNCGLNKAIPNLVLDHVAHGAAGRCQNALRYGVAL 173

QY 506 TNLTFGD--VANKATLCMSKGMALVAQLKSESEDLQOVIASVLRNLNWRADVNSKTL 563
DDB 174 TNLTFGDENVNKSILCQGRQFMVIAQLNAPDELLQVLQVLNLSWRADKMKMTIF 233

QY 564 REVGSVKALMECALEVKKESTLKVLSALWNLNLSAHCNTENKADICAVDGLAFVGLTYR 623
DDB 234 NELGTVTSLAARQNKNTLKALLSALWNLNLSAHCNTENKAEFCVADGLAFVGLNLSYE 293

QY 624 SQTNTLAIIESGGGILRNVSLLATNEDHRLNENCLQTLQHLKSHSITIVSNACGT 683
DDB 294 GPSKTLKIENAGGILRNVSIIAVCEPYRIILRYNCLAILLQQLKSESITIVSNSCGT 353

QY 684 LWNLSARNPKQDALWDMGAVSMLKNLIHSHKMIAMGSAALNLMANRPAXYKDNANIM 743
DDB 354 LWNLSARCPEDQOYLDDHNAIPILRALISSKNMIAEGSASALKNLVNFATBELMPN-- 411

QY 744 SPSSPSLHVRKOKALEAEADACHLSETPDNIDNLSPKASHRSKORHKOSLYGVYFDT 803
DDB 412 GDGSLP----- 418

QY 804 NRHDDNRSDNFNTGNMTVLSPLYNTTVLPSSSSRSGSLDSSRSRSEKORSLEEREGIGLNY 863
DDB 419 -----LDKEAGHG----- 426

QY 864 HPATENPGTSSKRGLOISTTAAQIAKVMEEVSAIHTSQEDRSSGSGTTELHCVTDERNALR 923
DDB 427 -----GTLPRR----- 432

QY 924 RSSAAHTHSNTYNTFKSENRTCSMPYAKLEYKRSNDLSNLSVSSSDGKGKQGMKPSI 983
DDB 433 -----FSSLRLSSNPTGSLKKVRFS-----TVSTTGFUNLRKSGRESI 470

QY 984 ESYSEDESKFCSYQGYPADLAHKIHSANHMDNDGELDTPINYSLKYSDEQLNSGRQSP 1043
DDB 471 YSGKSD-----STKYST----- 482

QY 1044 SQNERWAPKHIIEDIKQEQSRNQSTTYPVYTESTDDKHLKFPQHGQECVSPYR 1103
DDB 483 -----K 483

QY 1104 SRGANGSETNRVGNHGINQVSQLCQEDDYDDDKPTNYSERYSEEQHEEBERTNYS 1163
DDB 484 SEG-----KNPFIVTPT-----EQPIDYS 505

QY 1164 IKNEKRH-----VQPIDYSIKYATDIPSSQKQSFSSKSSGSSGSSKTEHMSSS 1215
DDB 506 MKTMEHKPNSSKTFEIDLQPTDFSARY-----KERRSAQTPKELASSETN 551

QY 1216 ENTSTPSSNAKRONQLHPSAQSR-----SQPOKAATCKVSSINQETIQYCVYE 1265
DDB 552 EIRS-----KELQTKSSATELNSPGLVAVSAKOKIAT-ETETETAEPINICEE 603

QY 1266 DTPICFRCSSLSLSLSAEDEIGCNOTQOASANTLQIAEIKETIGTRSAEDPVSEVA 1325
DDB 604 GTPGFSRFDLSNL-----TEKPE 623

QY 1326 VSQHPPTKSSRLQSGSLSSSARHKAVERFSSGNKSPSKGAQTPKSPPHYVQETPLMPS 1385
DDB 624 KCMPPKTPKT-----AVLEVHVDGNTPOIDS--ALETFMFS 660

QY 1386 RCTSVSLSDFSERSIA-----SSVQSEPCSGWVGIIISPSDLPDPSGQTMP--PSRSKTP 1439
DDB 661 RRSMDSLVG-DDETACEDNGSVISE-YRMOQSGVISELPSDPTQSMQSPRDR-- 716

QY 1440 PPPQTAQTKREVPKNAKPTAEKRESGPKQAAVAAVQVRVQLPDAATLLHFAFESTPDG 1499
DDB 717 ---KVSTQNNLDTPEOKPSTVE-----DKLNRPHVEHTPAA 750

QY 1500 FSCSSLSLSALDELPEFIQXDVLRIMPVQENDNGNETSEQPKESNCE-KAEKTD 1558
DDB 751 FSCATLSLSLSM-----MDSNANAIQGRDINGNDAPRSTCTED 793

QY 1559 SEKDLDDDDDDIEILEECIIISAMPTKSSRKAQPAQASKLPPVPA-----RKPSQ 1611
DDB 794 TTAVALSKAPNSDLSIL-----SIPNDLN-----ANEAQVPVAPRADVTGMDTMAPAE 841

QY 1612 LPVYKLLPSONRLOPKHVSFTPGDDMPVYCVGEGTPIFNSTATSLSDLTIESPPNELAA 1671
DDB 842 DAISKVRCGNALP-----SYLPVSDMSKYYVESPTFSVISGLSHUTV----- 887

QY 1672 GEGVRGAOSGFEKEDTTIPTGRSTDEAQGGKTSVTTPPELDNKK-ABEGDILAEICNS 1730
DDB 888 -----GSAKAGFLK---LPM--RTAEBAQA-----PKLPPRESAVCGD-----AEP 924

QY 1731 AMPKGSKHPFRVKKINDQ-----VQOASASSAPNKNQDLGKKKFTSPVKPIQONTYR 1786

```

Db      925  RLPPKSDLSLSSMSDDDDCNLLSCAIAAGSC-----RQPSGA--- 964
Qy      1787  TRVKWADSKNNLNAERFSDNKKDKQKLNKNSKDFDKLNNEDRVGSAFDSPHY 1846
Db      965  -----STSSSLANASTLCLRENGQSKQ-----VEHGDK-FN----- 996
Qy      1847  TPIEGTFCFSRNDLSLDFDDDDVDSREK-----AELRAKENKESEA----- 1892
Db      997  -----YSSDLSL-----DDDDDAKSLFEQCILSGHKSNDALSEGEPPGOR 1041
Qy      1893  -KVTSHTELTNSQ 1904
Db      1042  QEISARDRFVSNQ 1054

RESULT 13
Q961B0 PRELIMINARY; PRT; 1067 AA.
AC Q961B0;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE LB24920P
GN APC2 OR CG6193.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Eohydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.B., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051719; AAK93143.1; -
DR FlyBase; FBGN026598; Apc2.
DR InterPro; IPR008936; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 7.
DR PROSITE; PSS0176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116678 MW; 03C4119AEF19D198 CRC64;

Query Match      8.4%; Score 1221.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 9.7e-52;
Matches 429; Conservative 178; Mismatches 379; Indels 647; Gaps 50;

Qy      329  DMSRTLLAMSSQDSCTSMFQSCGLPLLIQLHGNKDKSVLLNGSRGSKAARASAAHL 388
Db      12  ELTRNLELRNPETCTALRSSDCIQLLVQLHANDE-----GLSTAKYASQALH 62
Qy      389  NIHSQDDKRGREIRVLHLEQIRAYCETC-WEWQEAHP-GMDQKNMPEAPVEHQI 446
Db      63  NIVHNPEEKREVRKMLDLLOILDYCNFLHTQJSGGEALADDEHRL----- 114
Qy      447  CPAYCVLMKLSFDEHHRAMNELLGGLQAIIEELQVDCENYG-LTNDHYSITLRRYAGMAL 505
Db      115  -AAMKLLMKASFDEHRQTMCELGAKAIPLNVLHDAVHGAAGMEQCNCALRSYGLMAL 173
Qy      506  TNLTFGD--YANKATLCSMKGCMEALVAOLKSESDIQQVIAVLRLNLSWRADVNSKTL 563
Db      174  TNLTFGDNVHNSKYLCCQGFQFMEVTAQNTAPDELLQVLQVLRNLSWRADNKHMTIF 233
Qy      564  REVGSVKALMECALVKKSTLKSVALNLSAHTENKADICAVDGAFLAVLGLTYR 623
Db      234  NELGTVTSLAAAMQNKNENTLKAILLSALNLSAHCSTNKAEPFCAVDGALAFVLGMLSYE 293
Qy      624  SQNTLAIISGGGILRNVSLSIATNEDHRLQILRENNCLTQLLOHLKSHSLTIVSNACGT 683

```

```

Db      294  GPSKTLKIIENAGGILRNVSISHAVCEPYQOILRYNCLAILLOQLKSESLTVNSCGT 353
Qy      684  LWNLSARNPKDQALWDMGAVSMKLNLIHSHKHMIAAGSAAALRNLMANRPAYKADANIM 743
Db      354  LWNLSARCPDQQLVDHNAIPLLRALISSKNSMIAEGSASALKVLNVPATBELMEN-- 411
Qy      744  SPGSSLSFLHVRKQKALEABLDQAHLSETPDNDNLNLPKASHRSKQKHOSLYGVYFDT 803
Db      412  GDGGLP----- 418
Qy      804  NRHDDNRSDNPTGNMTVLPYLNTTVLPSSSSSRGSLDSSRSSEKORSLEBRERIGLNY 863
Db      419  -----LDKEAGHG----- 426
Qy      864  HPATENPGTSKRGQLQISTTAAQIAKYMEEVSAIHTSQEDRSSGSTTELHCVTDERNALR 923
Db      427  -----GTLPRR----- 432
Qy      924  RSSAAHTSHNTYNTKSENSNRTCMPYAKLEYKRSNDSNLNSVSSDGGYKGGQMKPSI 983
Db      433  -----FSSLRSSNFTGSLKKVRPS-----TVSTTGFNLNRKCESRESI 470
Qy      984  ESYSEDDDESKFCSYGYQYPADLAHKAHSAHMDNDGELDTFINYSKYSDEQLNSCRQSP 1043
Db      471  YSGKSD-----STKYST----- 482
Qy      1044  SQNERWAPKHIIEDEIKQSEQRQSRNQSTTYPVYTESTDDKHLKFPQHFQGEQVSPYR 1103
Db      483  -----K 483
Qy      1104  SRGANGSETNRVSGNHGINQVNSQLCOEDDYEDDKPTNYSEYSEBEQHEEERPTNYS 1163
Db      484  SEGA-----KMPFEIVT-----BEQPIDYS 505
Qy      1164  IKYNEEKRH-----VDQPIDYSLKYATIDFSSQKQSFSSKSSGQSKSKTEHMSSS 1215
Db      506  MKYMEHFNSSKTFEIDLDTQDPSARY-----KERRSAQTAQPELKSETN 551
Qy      1216  ENTSTPSSNAKQNLHPSSNAQR-----SGOPKAACTCKVSSINQETIQTVCVE 1265
Db      552  EIRS-----KELQTKSSSATELRNSPGLVAUSAQKQIAT-ETETETAEPINCYEE 603
Qy      1266  DTPICFGRCSLSSLSAEDEIGCNQTTQEADSNLTQIAIEIKKICIGTSAEDPVSVPVA 1325
Db      604  GTPGSFSEFDSLSL-----TEKPE 623
Qy      1326  VSQHPRTKSSRLQSSLSSSSARHKAVEFSSGAKSPSKGAQTPKSPPEHVQVETPLMFS 1385
Db      624  KMPPTKPTKT-----AVLPVHVDGNTFQNIIDS--ALETPLMFS 660
Qy      1386  RCTSVSSLDSPESRSIA-----SSVQSPPCSGMWGIIISPDLPDPSPQQTWP--PSRSKTP 1439
Db      661  RRSMDSLVG--DDETACEDNGSVISB-YSRMQSGVISPSELDPSPQSMFQSPRRDR-- 716
Qy      1440  PPPQTAQTKREVKNKAPTAKRESGPKQAANVAQVRVQVLDPADTLHLHFAFESTPDG 1499
Db      717  ---KVSTQNNLDTPEQKPTVFE-----DKLARFHVHEHTPAA 750
Qy      1500  FSCSSLSALSLEDPFIQKDVLEIRMPVQNDNGNETESQPKESNENOB-KEAEKTIID 1558
Db      751  FSCATLSNLSM-----MDDSNANAIORGONDINGNGDAPSYCYTED 793
Qy      1559  SEKDILDDSDDDDEIIEECIISAMPTKSKSKAKKPAQTASKLPPVVA-----RPSQ 1611
Db      794  TTVLSKAPSNSDL-SIL-----SIPDLN-----ANEQFVAPRADVTDGMDTRPAE 841
Qy      1612  LPVYKLLPSQNLQPKQHVSTPGDDMPRVYCVGTPIINFSTATSLSDLTIESPPNELAA 1671
Db      842  DAISKORCGGNALP-----SYLPVSDENSKYVEDSPCTFSVISGLSHLTV----- 887
Qy      1672  GEGVGGAGQSSEFEKRDITPTTEGRSTDEAQGGKTSSVTIPELDNKA-ABEGDILAEICNS 1730

```

Db 888 -----GSAKAGPVLK---LPM--RTAEAAQA-----PKLPPRSVQGD-----AEP 924

QY 1731 AMPKGSHKPPRVKIMDO-----VOASASSAPNKNQDGGKKKPTGPVKIPQNTBYR 1786

Db 925 RLFPKSDLSLSLSDSDCCNLLSQIAAGSC-----RPQPSGA--- 964

QY 1787 TVRVKNADSKNLNLAERVPFSDKSKONLKNNSXDENDKLPNNEDVRGSPAFDPSPHY 1846

Db 965 -----STSSSLANASTILCRENGQSKQ-----VEHGDK-PN----- 996

QY 1847 TPIETPYCFNRNDSLSLDFDDVDLSREK-----AELRKAKENKSEA----- 1892

Db 997 -----YSSDDSL-----DDDDARSLSLFEQILSGMKNSDALSSEGBPPQQR 1041

QY 1893 -KVTSHTELTNSQ 1904

Db 1042 QEISARDRFVSNQ 1054

RESULT 14

Q9P119

ID Q9P119 PRELIMINARY; PRT; 208 AA.

AC Q9P119

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Adenomatosis polyposis coli tumor suppressor (fragment).

GN APC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20435068; PubMed=10982189;

RA Su L.K., Steinbach G., Sawyer J.C., Hindi M., Ward P.A., Lynch P.M.;

RT "Genomic rearrangements of the APC tumor-suppressor gene in familial

RT adenomatous polyposis";

RL Hum. Genet. 106:101-107(2000).

DR EMBL; AF127506; AAF34355.1; .

DR EMBL; AF127034; AAF34355.1; JOINED.

DR InterPro; IPR008938; ARM.

DR InterPro; IPR000225; Armadillo.

DR Pfam; PF00514; Armadillo_seg; 2.

FT NON_TER 1

FT NON_TER 208

SQ SEQUENCE 208 AA; 22726 MW; 51076F257B61C528 CRC64;

Query Match 7.1%; Score 1041; DB 4; Length 208;

Best Local Similarity 100.0%; Pred. No. 8.1e-44;

Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 GLQAIAEQLQVDCMYGLTNDHYSITLRRYAGMALTNLTFGDVANKATLCSMKGCMRALV 530

Db 1 GLQAIAEQLQVDCMYGLTNDHYSITLRRYAGMALTNLTFGDVANKATLCSMKGCMRALV 60

QY 531 AQLKSESEDLQVIVASVLRLNLSRADVNSKTLREVGSKALMECALEVKKESTLKSVL 590

Db 61 AQLKSESEDLQVIVASVLRLNLSRADVNSKTLREVGSKALMECALEVKKESTLKSVL 120

QY 591 ALWNLSAHCENKADICAVDAGALFLVGLTLYRSQTNLTALIESGGGILRLNVSSLIATNE 650

Db 121 ALWNLSAHCENKADICAVDAGALFLVGLTLYRSQTNLTALIESGGGILRLNVSSLIATNE 180

QY 651 DHRQILRENCLQTLLOHLKSHSLTIVS 678

Db 181 DHRQILRENCLQTLLOHLKSHSLTIVS 208

RESULT 15

Q7Z2Q8

ID Q7Z2Q8 PRELIMINARY; PRT; 159 AA.

AC Q7Z2Q8

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein (fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932;

RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Strausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RX Strausberg R.;

RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC056268; AAHS6268.1; .

KW Hypothetical protein.

FT NON_TER 159

FT NON_TER 159

SQ SEQUENCE 159 AA; 17843 MW; E0C0CC055A22C91B CRC64;

Query Match 5.2%; Score 756; DB 4; Length 159;

Best Local Similarity 98.7%; Pred. No. 5.9e-30;

Matches 153; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAASYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKVKLVKLOQGSIDEAM 60

Db 1 MAAASYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKVKLVKLOQGSIDEAM 60

QY 61 ASSGQIDLLERLKEMLNDSNFPFGVKLRKMSLRYSYSGREGSVSSRSGCSPVPMGSPFR 120

Db 61 ASSGQIDLLERLKEMLNDSNFPFGVKLRKMSLRYSYSGREGSVSSRSGCSPVPMGSPFR 120

QY 121 RGFVNGSRSTGYLELEKERSLLADLKEKEK 155

Db 121 RGFVNGSRSTGYLELEKERSLLADLKEKEK 155

Search completed: August 25, 2004, 17:23:07

Job time : 214.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:01 ; Search time 180.5 Seconds

(without alignments)
4450.319 Million cell updates/sec

Title: US-09-442-489f-7

Perfect score: 14566

Sequence: 1 MAASVDQLQLKVEALKMEN.....ESSGTQPKRHSGLVLTSS 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14566	100.0	2843	2 AAW76140	AAW76140 Human APC
2	14566	100.0	2843	2 AAW76144	AAW76144 Human APC
3	14566	100.0	2843	3 AAB23011	AAB23011 Human APC
4	14566	100.0	2843	5 ABG71105	ABG71105 Human ade
5	14566	100.0	2973	2 AAW76821	AAW76821 Human APC
6	14566	100.0	2973	4 AAY72782	AAY72782 Transcrip
7	14559	100.0	2973	3 AAY70304	AAY70304 Protein u
8	14548.5	99.9	2842	2 AAR63508	AAR63508 Adenomat
9	14548.5	99.9	2842	5 AAG90968	AAG90968 Human APC
10	14546	99.9	2843	2 AAW11922	AAW11922 Adenomat
11	14539	99.8	2843	7 ADE65846	ADE65846 Human ade
12	14533	99.8	2843	2 AAR26052	AAR26052 APC gene
13	14533	99.8	2843	2 AAW35392	AAW35392 Human ade
14	14533	99.8	2843	2 AAW38370	AAW38370 Human ade
15	14533	99.8	2843	5 ABG90964	ABG90964 Human ade
16	14509	99.6	2860	2 AAR63507	AAR63507 Adenomat
17	14479	99.4	2843	2 AAR58634	AAR58634 Adenomat
18	13999.5	96.1	2742	3 AAB23012	AAB23012 Human APC
19	13198.5	90.6	2842	7 ADE56175	ADE56175 Rat Prote
20	4738	32.5	912	5 ABG71106	ABG71106 Human ade
21	3892	26.7	767	5 ABG71107	ABG71107 Human ade
22	3512.5	24.1	2274	4 AAB50574	AAB50574 Mouse APC
23	3488.5	23.9	902	4 AAG09335	AAG09335 Novel hum
24	3415.5	23.4	2303	6 ABR58648	ABR58648 Human can
25	2169.5	14.9	799	3 AAY92061	AAY92061 Human APC

ALIGNMENTS

RESULT 1

AAW76140
ID AAW76140 standard; protein; 2843 AA.

XX AC AAW76140;

XX DT 25-MAR-2003 (revised)

XX DT 23-NOV-1998 (first entry)

XX DE Human APC protein #1.

XX KW Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;

XX KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;

XX KW Gardner's Syndrome; GS; predisposition.

XX OS Homo sapiens.

XX PN US5783666-A.

XX PD 21-JUL-1998.

XX PF 25-MAY-1995; 95US-00452655.

XX PR 16-JAN-1991; 91GB-00000962.

XX PR 16-JAN-1991; 91GB-00000963.

XX PR 16-JAN-1991; 91GB-00000974.

XX PR 16-JAN-1991; 91GB-00000975.

XX PR 08-AUG-1991; 91US-00741940.

XX PR 12-AUG-1994; 94US-00289548.

XX PA (CANC-) CANCER INST.

XX PA (UYUO) UNIV JOHNS HOPKINS.

XX PA (UTAH) UNIV UTAH.

XX PA (ZENE) ZENECA PHARM.

XX PI Kinzler K, Joslyn G, Markham AP, Carlson M, White RL;

XX PI Thliveris A, Groden J, Anand R, Nakamura Y, Albertsen H;

XX PI Vogelstein B, Hedge PJ;

XX DR WPI; 1998-427100/36.

XX DR N-PSDB; AAV56447.

XX PT Adenomatous polyposis coli protein - useful in the treatment of cancers

XX PT associated with mutation(s) on human chromosome 5q21.

XX PS Disclosure; Col 41-54; 102pp; English.

XX CC This sequence represents a human familial adenomatous polyposis coli

AY92060 Murine AP
AB58126 Drosophil
AAR8353 Mutant A-
AB62156 Drosophil
AAU37120 Staphyloc
AB68397 Drosophil
AB67502 Drosophil
AB66878 Drosophil
ABU16000 Protein e
ABM72734 Staphyloc
ABP56876 Staphyloc
ABJ18914 Pathogen
AB50654 C. elegans
AAW33894 Flea sali
AAW82368 Flea sali
ABP39618 Staphyloc
AB60327 Drosophil
AB60291 Drosophil
AB65772 Drosophil
AB671160 Drosophil

CC (APC) protein from clone pp2.5. The gene for the protein is present on
 CC human chromosome 5q21 and is also referred to as adenomatous polyposis
 CC coli gene. It is a tumour suppressor gene, and mutations in this gene
 CC have been associated with tumorigenesis in retinoblastoma and colorectal
 CC tumours, and especially familial adenomatous polyposis (FAP) and
 CC Gardner's Syndrome (GS). The protein can be used in therapy to replace
 CC lack of native functional protein and the nucleic acids can be used for
 CC gene therapy. The nucleic acids that encode them can also be used as
 CC probes and primers in detection of the cancers and predisposition to it.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC
 XX
 SQ

Sequence 2843 AA;

Query Match 100.0%; Score 14566; DB 2; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAASDYDQLKQVEALKMNSNLRQLEDSNHLTKLETEASNKKEVLKQLOGSIEDEAM	60
DB	1	MAAASDYDQLKQVEALKMNSNLRQLEDSNHLTKLETEASNKKEVLKQLOGSIEDEAM	60
QY	61	ASSQIDLLRLKELNLDSSNFFGVKLRKMSLRSGREGSVSSRSGECSPVPMGSPFR	120
DB	61	ASSQIDLLRLKELNLDSSNFFGVKLRKMSLRSGREGSVSSRSGECSPVPMGSPFR	120
QY	121	RGFVNGSRESTGYLEBELEKERSILLADLDKEEKEKDWYQAOLNLTKRIDSPLTENFSL	180
DB	121	RGFVNGSRESTGYLEBELEKERSILLADLDKEEKEKDWYQAOLNLTKRIDSPLTENFSL	180
QY	181	QDWTNRQLYEARQIRVAMEEQLGTQCDMEKKAQRRIARIQIIEKDILRIQLQSQAT	240
DB	181	QDWTNRQLYEARQIRVAMEEQLGTQCDMEKKAQRRIARIQIIEKDILRIQLQSQAT	240
QY	241	EAERSSQNHETGSHDAERONEGGVGEINMATSGNGQGSQTRMDHETASVLSSSSTHSA	300
DB	241	EAERSSQNHETGSHDAERONEGGVGEINMATSGNGQGSQTRMDHETASVLSSSSTHSA	300
QY	301	PRRLTSLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCIEMROSGCLPLLIQLL	360
DB	301	PRRLTSLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCIEMROSGCLPLLIQLL	360
QY	361	HGNDKOSVLLNGSRGSEKARASAAALHNIHSDPDQKRRREIRVLHLEQIRAYCETC	420
DB	361	HGNDKOSVLLNGSRGSEKARASAAALHNIHSDPDQKRRREIRVLHLEQIRAYCETC	420
QY	421	WENQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQIAIELLO	480
DB	421	WENQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQIAIELLO	480
QY	481	VDCEMYGLTNDHYSITLRRYAGKALTNLTFGDVANKATLCSMGKCMRVALVAQLKSEEDL	540
DB	481	VDCEMYGLTNDHYSITLRRYAGKALTNLTFGDVANKATLCSMGKCMRVALVAQLKSEEDL	540
QY	541	QOVIASVLRNLWRADVNSKTLREVGSVKALMECALEVKKESTLKSVALNLSAHT	600
DB	541	QOVIASVLRNLWRADVNSKTLREVGSVKALMECALEVKKESTLKSVALNLSAHT	600
QY	601	ENKADICAVDGAFLVGTITYSQNTLAIIESGGGILRNVSLLATNEDHRQLRENN	660
DB	601	ENKADICAVDGAFLVGTITYSQNTLAIIESGGGILRNVSLLATNEDHRQLRENN	660
QY	661	CLQTLLOHLKSHSLTIVSNACGTLWNLSARNPKDQALWDMGAVSMKLIHSHKHVAM	720
DB	661	CLQTLLOHLKSHSLTIVSNACGTLWNLSARNPKDQALWDMGAVSMKLIHSHKHVAM	720
QY	721	GSAALRNLMANRPAAKNDANIMSPGSSLPFLVRKQKALEAEALDAQHLSSETFDNIDNLS	780
DB	721	GSAALRNLMANRPAAKNDANIMSPGSSLPFLVRKQKALEAEALDAQHLSSETFDNIDNLS	780
QY	781	PKASHRSKORHKOSLYGDYVFTNRHDDNRSDFNTGNMTVLPSPYNTTVPSSSSSRGS	840
DB	781	PKASHRSKORHKOSLYGDYVFTNRHDDNRSDFNTGNMTVLPSPYNTTVPSSSSSRGS	840

QY	841	LDSSRSRSEKRSLESERGIGLGNVHPATENPGTSSKRGLOISTTAAQIAKVNBEVSAIHTS	900
DB	841	LDSSRSRSEKRSLESERGIGLGNVHPATENPGTSSKRGLOISTTAAQIAKVNBEVSAIHTS	900
QY	901	QEDRSSSTTELHCVTDERNALRESSAAHNTNFTKSENSNRTCSMPYAKLEYKRSS	960
DB	901	QEDRSSSTTELHCVTDERNALRESSAAHNTNFTKSENSNRTCSMPYAKLEYKRSS	960
QY	961	NDSLNSVSSSDGYKRGQMKPSTIESYSEDDSKPCSYQYPADLAHKIHSANHMDNDGE	1020
DB	961	NDSLNSVSSSDGYKRGQMKPSTIESYSEDDSKPCSYQYPADLAHKIHSANHMDNDGE	1020
QY	1021	LDTPINTSLKYSDROLNSGRSQSQNERWARPKHIIIDEIKQSEORQSRNQSTPYVTE	1080
DB	1021	LDTPINTSLKYSDROLNSGRSQSQNERWARPKHIIIDEIKQSEORQSRNQSTPYVTE	1080
QY	1081	STDDKHLKFPHFQOQECVSPYRSRGANGSETNRVGNHGINQNVQSCLQEDDYEDDKP	1140
DB	1081	STDDKHLKFPHFQOQECVSPYRSRGANGSETNRVGNHGINQNVQSCLQEDDYEDDKP	1140
QY	1141	TNYSERYSSEHQHEBERPTNYSIKYNEKRHVDPIDYSLKYATDIPSSQKQSFSEKS	1200
DB	1141	TNYSERYSSEHQHEBERPTNYSIKYNEKRHVDPIDYSLKYATDIPSSQKQSFSEKS	1200
QY	1201	SSGSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQSTIQ	1260
DB	1201	SSGSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQSTIQ	1260
QY	1261	TYCVEDTPICFSRCSLSSIAEDEIGCMQTTQEDASANTLQIAEIKKIGTRSAEDPV	1320
DB	1261	TYCVEDTPICFSRCSLSSIAEDEIGCMQTTQEDASANTLQIAEIKKIGTRSAEDPV	1320
QY	1321	SEVFAVSQHPHTKSRLOQSSLSSESARHKAVERSSGAKSPSKGAOTPKSPPHYVQET	1380
DB	1321	SEVFAVSQHPHTKSRLOQSSLSSESARHKAVERSSGAKSPSKGAOTPKSPPHYVQET	1380
QY	1381	PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVGIIISPSDLPDPSGQTMPPSRSTPP	1440
DB	1381	PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVGIIISPSDLPDPSGQTMPPSRSTPP	1440
QY	1441	PPQTAQTKRVPKNAKTAETAKRESGPKQAVNAQVQVLPDADTLHFAFTESTPDGF	1500
DB	1441	PPQTAQTKRVPKNAKTAETAKRESGPKQAVNAQVQVLPDADTLHFAFTESTPDGF	1500
QY	1501	SCSSLSALSALDEPFIQKVELRIMPVQENDNGNETESQPKESNENQEKAEKTTDSE	1560
DB	1501	SCSSLSALSALDEPFIQKVELRIMPVQENDNGNETESQPKESNENQEKAEKTTDSE	1560
QY	1561	KOLLDDSDDDIIELEECIISAMPTKSRKAKPAQATASKLPPPVARKPQOLPYKLLPS	1620
DB	1561	KOLLDDSDDDIIELEECIISAMPTKSRKAKPAQATASKLPPPVARKPQOLPYKLLPS	1620
QY	1621	QNRLOPKHVSFTPDGDMRPVYCVVEGTPINFSTATISLSDLTIESPPNELAAGEGVGGAQ	1680
DB	1621	QNRLOPKHVSFTPDGDMRPVYCVVEGTPINFSTATISLSDLTIESPPNELAAGEGVGGAQ	1680
QY	1681	SGEFKERTIIEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEICINSAMPKSKHP	1740
DB	1681	SGEFKERTIIEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEICINSAMPKSKHP	1740
QY	1741	FRVKKIMDOVOQASASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNKL	1800
DB	1741	FRVKKIMDOVOQASASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNKL	1800
QY	1801	AERFSDNCKSKQNLKNNSKDFNDKLPNNEDVRGSAFDSPHHYTPIEGTPTCYFGRND	1860
DB	1801	AERFSDNCKSKQNLKNNSKDFNDKLPNNEDVRGSAFDSPHHYTPIEGTPTCYFGRND	1860
QY	1861	SLSLDPPDDDDVLSREKAELEKAKENKESPAKVTSHTELTSNOOSANKTQAIKQPINR	1920
DB	1861	SLSLDPPDDDDVLSREKAELEKAKENKESPAKVTSHTELTSNOOSANKTQAIKQPINR	1920
QY	1921	QQPKPILOKSTFFQSSKDIIPRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN	1980

Db 1921 GPKPIQKSTFPQSSKIDPRGAATDEKLNQFAIENTPVCFNSHSLSDIDQNN 1980
 QY 1981 NKENEPIKETEPPDSGEGSKQASGYAPKSPHVEDTPVCFNSHSLSDIDQNN 2040
 Db 1981 NKENEPIKETEPPDSGEGSKQASGYAPKSPHVEDTPVCFNSHSLSDIDQNN 2040
 QY 2041 ECISAMPKXKXKPSRLKGNKHSNNMGIIIGEDLTLDKDIQRPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKXKXKPSRLKGNKHSNNMGIIIGEDLTLDKDIQRPDSEHGLSPDSENF 2100
 QY 2101 WKAIQEGANSIVSSLHQAACLSROASSDSLSLSKSGISLSPHLPDQEKPT 2160
 Db 2101 WKAIQEGANSIVSSLHQAACLSROASSDSLSLSKSGISLSPHLPDQEKPT 2160
 QY 2161 SNKGRPIILKPGKSTLETKKIESEKGIKGGKVKVKSITGKVRNSISGOMKQLOAN 2220
 Db 2161 SNKGRPIILKPGKSTLETKKIESEKGIKGGKVKVKSITGKVRNSISGOMKQLOAN 2220
 QY 2221 MFSISGRGTMIHIGVRNNSSTSPVSKGPPKTPASKPSGEGTATTSRGAKPVS 2280
 Db 2221 MFSISGRGTMIHIGVRNNSSTSPVSKGPPKTPASKPSGEGTATTSRGAKPVS 2280
 QY 2281 ELSPVARQTSQIGGSKAPSRGSDSTPSRPAQOPLSRPIQSPGRNISPGRNGISPPN 2340
 Db 2281 ELSPVARQTSQIGGSKAPSRGSDSTPSRPAQOPLSRPIQSPGRNISPGRNGISPPN 2340
 QY 2341 KLSQLPRTSSPTASTKSSGSKMYSVTPGRQMSQNLTKQTLGSKNASSIPRSESASK 2400
 Db 2341 KLSQLPRTSSPTASTKSSGSKMYSVTPGRQMSQNLTKQTLGSKNASSIPRSESASK 2400
 QY 2401 LNMNNGNGANKVELSRMSSVSKSGSDRSPVLRQSTFIKEAPSPTLRKLSSA 2460
 Db 2401 LNMNNGNGANKVELSRMSSVSKSGSDRSPVLRQSTFIKEAPSPTLRKLSSA 2460
 QY 2461 SFESLSPSSRPASPTRSQATPVLSPSLPDMSLSTHSSVQAGGWRKLPNLSPTIYNDG 2520
 Db 2461 SFESLSPSSRPASPTRSQATPVLSPSLPDMSLSTHSSVQAGGWRKLPNLSPTIYNDG 2520
 QY 2521 RPAKRDIAHSESPRLPINSRTGWRKHSKSSSLPRVSTWRTGSSSILSASSES 2580
 Db 2521 RPAKRDIAHSESPRLPINSRTGWRKHSKSSSLPRVSTWRTGSSSILSASSES 2580
 QY 2581 SEKASEDEKHVNSISGTSKQENQVSAKGTWRKIKENEFPTNSTSTVSSGATNGAES 2640
 Db 2581 SEKASEDEKHVNSISGTSKQENQVSAKGTWRKIKENEFPTNSTSTVSSGATNGAES 2640
 QY 2641 KTLIIQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTNPVIDSVSKANPNIKDSKN 2700
 Db 2641 KTLIIQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTNPVIDSVSKANPNIKDSKN 2700
 QY 2701 QAKQVNGSVPMRTVGLNRLNSFIQVADPQKTEIKPGQNNPVVSETNESSIVERT 2760
 Db 2701 QAKQVNGSVPMRTVGLNRLNSFIQVADPQKTEIKPGQNNPVVSETNESSIVERT 2760
 QY 2761 PFSSSSSKHSSPSGTVAARVTPFNPNPSPRKSSADTSARPSQIPTPVNNTKKRDSKT 2820
 Db 2761 PFSSSSSKHSSPSGTVAARVTPFNPNPSPRKSSADTSARPSQIPTPVNNTKKRDSKT 2820
 QY 2821 DSTESSGTQSPKXHSGLVTSV 2843
 Db 2821 DSTESSGTQSPKXHSGLVTSV 2843

RESULT 2
 AAW76144
 ID AAW76144 standard; protein; 2843 AA.
 XX
 AC AAW76144;
 XX
 DT 25-MAR-2003 (revised)
 DT 23-NOV-1998 (first entry)
 XX

DE XX Human APC protein #2.
 KW Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
 KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
 KW Gardner's Syndrome; GS; predisposition.
 XX
 OS Homo sapiens.
 XX
 PN US5783666-A.
 XX
 PD 21-JUL-1998.
 XX
 PF 25-MAY-1995; 95US-00452655.
 XX
 PR 16-JAN-1991; 91GB-00000962.
 PR 16-JAN-1991; 91GB-00000963.
 PR 16-JAN-1991; 91GB-00000974.
 PR 16-JAN-1991; 91GB-00000975.
 PR 08-AUG-1991; 91US-00741940.
 PR 12-AUG-1991; 94US-00289548.
 XX
 PA (CANC-) CANCER INST.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UTAH) UNIV UTAH.
 PA (ZENE) ZENECA PHARM.
 XX
 PI Kinzler K, Joslyn G, Markham AP, Carlson M, White RL;
 PI Thliveris A, Groden J, Anand R, Nakamura Y, Albertsen H;
 PI Vogelstein B, Hedge PJ;
 XX
 DR WPI; 1998-427100/36.
 XX
 PT Adenomatous polyposis coli protein - useful in the treatment of cancers
 PT associated with mutation(s) on human chromosome 5q21.
 XX
 PS Disclosure; Col 63-78; 102pp; English.
 CC This sequence represents a human familial adenomatous polyposis coli
 CC (APC) protein isolated from 87 cDNA clones. The gene for the protein is
 CC present on human chromosome 5q21 and is also referred to as adenomatous
 CC polyposis coli gene. It is a tumour suppressor gene, and mutations in
 CC this gene have been associated with tumorigenesis in retinoblastoma and
 CC colorectal tumours, and especially familial adenomatous polyposis (FAP)
 CC and Gardner's Syndrome (GS). The protein can be used in therapy to
 CC replace lack of native functional protein and the nucleic acids can be
 CC used for gene therapy. The nucleic acids that encode them can also be
 CC used as probes and primers in detection of the cancers and predisposition
 CC to it. (Updated on 25-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 2843 AA;
 Query Match 100.0%; Score 14566; DB 2; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAASYDQLLKQVEALKWENSLRQLEDNSNHLTKLETEASNKVEYLKQGSIEDEAM 60
 Db 1 MAAASYDQLLKQVEALKWENSLRQLEDNSNHLTKLETEASNKVEYLKQGSIEDEAM 60
 QY 61 ASSQIDLLRLKELNLDSSNPFQVKLRSMVRSYSGREGSVSSRSGECPVPMGSPFR 120
 Db 61 ASSQIDLLRLKELNLDSSNPFQVKLRSMVRSYSGREGSVSSRSGECPVPMGSPFR 120
 QY 121 RGFVNGSRESTGYLEELKERSLLDLADLKEEKEKWYVLAQLNLTKRIDSLPTNFSL 180
 Db 121 RGFVNGSRESTGYLEELKERSLLDLADLKEEKEKWYVLAQLNLTKRIDSLPTNFSL 180
 QY 181 QTMTRQLEYEARIQIRVAMEEQIGTCQDMKEKQAQRRIARIQIETKDIIRIQLQSQA 240
 Db 181 QTMTRQLEYEARIQIRVAMEEQIGTCQDMKEKQAQRRIARIQIETKDIIRIQLQSQA 240
 QY 241 EABESSONKHETGSHDAERONEGOGVEINWATSGNGQSTTMDHETASVSSSTHSA 300

Db 241 EAERSSQNKHETGSHDAERQNEGQGVCEINMATSNGQGGSTTMDHETASVLSSSTHSA 300
Qy 301 PRRLTSHLGTKEVMVYSLLSMLGTHDKDDMSRRTLWAMSSQDSCIWMROSGLPLLIQLL 360
Db 301 PRRLTSHLGTKEVMVYSLLSMLGTHDKDDMSRRTLWAMSSQDSCIWMROSGLPLLIQLL 360
Qy 361 HGKDKSVLGNRSGSEKARASAAALHNLIHSQDDKGRREIRVLHLEQIRAYCETC 420
Db 361 HGKDKSVLGNRSGSEKARASAAALHNLIHSQDDKGRREIRVLHLEQIRAYCETC 420
Qy 421 WEMQEAHEPGMDQKNPMPAPVHQICPAVCVLMKLSFDEHRHAMNELGGLQAI AELLQ 480
Db 421 WEMQEAHEPGMDQKNPMPAPVHQICPAVCVLMKLSFDEHRHAMNELGGLQAI AELLQ 480
Qy 481 VDCBMYGLTNDHYSITLRRVAGMALTNLTGDIVANKATLCSMKGCMRALVAQLKSEEDL 540
Db 481 VDCBMYGLTNDHYSITLRRVAGMALTNLTGDIVANKATLCSMKGCMRALVAQLKSEEDL 540
Qy 541 QOVLASVLRNLNWRADVNSKKTLEVGSVKALMECALEVKKESTLXSVLSALWNL SAHCT 600
Db 541 QOVLASVLRNLNWRADVNSKKTLEVGSVKALMECALEVKKESTLXSVLSALWNL SAHCT 600
Qy 601 ENKADI CAVDGAALFLVGTITYRSQTNLTALIESGGGILRNVSLLATNEDHROI LRENN 660
Db 601 ENKADI CAVDGAALFLVGTITYRSQTNLTALIESGGGILRNVSLLATNEDHROI LRENN 660
Qy 661 CLOTLLOHLKSHSLTIVSNACGTLWNLARNPKDQALWDMGAVSMLKNLIHSKHWMIAM 720
Db 661 CLOTLLOHLKSHSLTIVSNACGTLWNLARNPKDQALWDMGAVSMLKNLIHSKHWMIAM 720
Qy 721 GSAAALRNLMANRPKADANIMSPGSLPSLHV RKQKALEADAOHLSETFDNIDNLS 780
Db 721 GSAAALRNLMANRPKADANIMSPGSLPSLHV RKQKALEADAOHLSETFDNIDNLS 780
Qy 781 PKASHRSKQRHKSQSLYGDYVFTNRHDDNDSDFNTGNMTVLSPYLTNTVLPSSSSRGS 840
Db 781 PKASHRSKQRHKSQSLYGDYVFTNRHDDNDSDFNTGNMTVLSPYLTNTVLPSSSSRGS 840
Qy 841 LDSRSSEKDRSLERERIGLGNHYHPATENFCTSSKRGLOISTTAAQIAKMEVSAIHTS 900
Db 841 LDSRSSEKDRSLERERIGLGNHYHPATENFCTSSKRGLOISTTAAQIAKMEVSAIHTS 900
Qy 901 QEDRSSGSTTELHCVTDERNALRSSAAHSTNFTNFKSENSNRITCSMPYAKLEYKRSS 960
Db 901 QEDRSSGSTTELHCVTDERNALRSSAAHSTNFTNFKSENSNRITCSMPYAKLEYKRSS 960
Qy 961 NDSLNSVSSSDGYKRGOMKPSIESYEDDESKEFCSYGQVPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLNSVSSSDGYKRGOMKPSIESYEDDESKEFCSYGQVPADLAHKIHSANHMDNDGE 1020
Qy 1021 LDTPTINTSLKYSDQLNSGROSPQNERWARPKHII EDEIKOSEQRQSRNQSTTYPVYTE 1080
Db 1021 LDTPTINTSLKYSDQLNSGROSPQNERWARPKHII EDEIKOSEQRQSRNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFQPHFGQECVSPYRSGRANGSETNRVGNHGINQVNSQSLCEQDDYEDDKP 1140
Db 1081 STDDKHLKFQPHFGQECVSPYRSGRANGSETNRVGNHGINQVNSQSLCEQDDYEDDKP 1140
Qy 1141 TNYERYSEEPQHEERPTNYSIKYNEKRVHDQPIDYSLKYATDIPSSQKQSFPSKSS 1200
Db 1141 TNYERYSEEPQHEERPTNYSIKYNEKRVHDQPIDYSLKYATDIPSSQKQSFPSKSS 1200
Qy 1201 SSGOSSKTEHMSSENSTSPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQBTIQ 1260
Db 1201 SSGOSSKTEHMSSENSTSPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQBTIQ 1260
Qy 1261 TYCVEDTPICFSRCSLSSLSABDEIGCNQTTQEADSANTLQIAIEKEXIGTRSABDPV 1320
Db 1261 TYCVEDTPICFSRCSLSSLSABDEIGCNQTTQEADSANTLQIAIEKEXIGTRSABDPV 1320
Qy 1321 SEVPAVSQHPRTKSRLOGSSLSESARHKAVESSGAKSPSKGAOTPKSPPEHYVQET 1380
Db 1321 SEVPAVSQHPRTKSRLOGSSLSESARHKAVESSGAKSPSKGAOTPKSPPEHYVQET 1380

Qy 1381 PLMFBSRCTSVSSLDSPBSRSRTASSVQSPGCMVSGIISPSDLPDSPGQTMPPRSKTI PP 1440
Db 1381 PLMFBSRCTSVSSLDSPBSRSRTASSVQSPGCMVSGIISPSDLPDSPGQTMPPRSKTI PP 1440
Qy 1441 PPPQTAQTKREVPKNKAAPTAEKRESGPKQAANAQVQVLPDADTLLHFATESTPDGF 1500
Db 1441 PPPQTAQTKREVPKNKAAPTAEKRESGPKQAANAQVQVLPDADTLLHFATESTPDGF 1500
Qy 1501 SCSSLSALSALDEBFIQKVELRIMPVQENDNGNETESQPKESNENQKAEKTI DSE 1560
Db 1501 SCSSLSALSALDEBFIQKVELRIMPVQENDNGNETESQPKESNENQKAEKTI DSE 1560
Qy 1561 KOLLDDSDDDDIETLEECIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KOLLDDSDDDDIETLEECIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATISLTLTIESPPNELAAGEVRGAQ 1680
Db 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATISLTLTIESPPNELAAGEVRGAQ 1680
Qy 1681 SCEPEKRTIIPTEGRSTDEAOGKTSVVTIPELDDNKAEBGDI LAECINSAMPKKGSHKP 1740
Db 1681 SCEPEKRTIIPTEGRSTDEAOGKTSVVTIPELDDNKAEBGDI LAECINSAMPKKGSHKP 1740
Qy 1741 FRVKKIMDQVOQAASASSAPNKNQIDGKKKXPTSPVXPIQONTYRTRVRKNADSKN LN 1800
Db 1741 FRVKKIMDQVOQAASASSAPNKNQIDGKKKXPTSPVXPIQONTYRTRVRKNADSKN LN 1800
Qy 1801 AERFESDNKDKSKNKNKNNDKNDKLPNNEDRVRGSPAFDPSPHYPIETGTYCFSRND 1860
Db 1801 AERFESDNKDKSKNKNKNNDKNDKLPNNEDRVRGSPAFDPSPHYPIETGTYCFSRND 1860
Qy 1861 SLSSLDLDDDDVDVLSREKAELRKAKENKESBAKVTSHTELTSNQOSANKTOAI AKQPINR 1920
Db 1861 SLSSLDLDDDDVDVLSREKAELRKAKENKESBAKVTSHTELTSNQOSANKTOAI AKQPINR 1920
Qy 1921 GQPKILOQKSTFQSSKDI PDGNAATDEKLONFAIENTPVCFSHNSLSLSLSDIDENN 1980
Db 1921 GQPKILOQKSTFQSSKDI PDGNAATDEKLONFAIENTPVCFSHNSLSLSLSDIDENN 1980
Qy 1981 NKENEPIKETETPPDPSQGEPSKPQAGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDL LQ 2040
Db 1981 NKENEPIKETETPPDPSQGEPSKPQAGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDL LQ 2040
Qy 2041 ECISAMPKKKPSRLKGDNEKXSPRNMGGLGEDLTLDLKDIORPDEHGLSPDSNFD 2100
Db 2041 ECISAMPKKKPSRLKGDNEKXSPRNMGGLGEDLTLDLKDIORPDEHGLSPDSNFD 2100
Qy 2101 WKAIQEGANSIVSSLHQAAAAACL SRQASSDSDSLKSGISLCSPPHLLTPDOEKPFT 2160
Db 2101 WKAIQEGANSIVSSLHQAAAAACL SRQASSDSDSLKSGISLCSPPHLLTPDOEKPFT 2160
Qy 2161 SNKGPRILLKPGEKSTLETTKIIESSKGIKGGKVKYKSLITGKVRNSBI SQMKQPLQAN 2220
Db 2161 SNKGPRILLKPGEKSTLETTKIIESSKGIKGGKVKYKSLITGKVRNSBI SQMKQPLQAN 2220
Qy 2221 MPISI SRGRTMIHI PGVRNSSSTSPVSKKGPPLKTPASKSPSEGTATTSRGA KPSVK 2280
Db 2221 MPISI SRGRTMIHI PGVRNSSSTSPVSKKGPPLKTPASKSPSEGTATTSRGA KPSVK 2280
Qy 2281 ELSPVARQTSQIGGSSKAPSRGSRDSTPSRPAQPLSRPIQSPGRNISIPGRNGI SPPN 2340
Db 2281 ELSPVARQTSQIGGSSKAPSRGSRDSTPSRPAQPLSRPIQSPGRNISIPGRNGI SPPN 2340
Qy 2341 KLSQLPRTSSPSTASTKSSGSKWYTPGQMSQONLTQOTGLSKNASSI PRSESASKG 2400
Db 2341 KLSQLPRTSSPSTASTKSSGSKWYTPGQMSQONLTQOTGLSKNASSI PRSESASKG 2400
Qy 2401 LNMQNGANGANKVELSRMSSTKSSGESDRSERPVLVRQSTFIKEAPSTLRRLKLESA 2460
Db 2401 LNMQNGANGANKVELSRMSSTKSSGESDRSERPVLVRQSTFIKEAPSTLRRLKLESA 2460

QY 2461 SFESLSPSPASPSPTRSQATPVLSPSLPDMSLSTHSSVQAGWKLPPNLSPPTIEYNDG 2520
Db 2461 SFESLSPSPASPSPTRSQATPVLSPSLPDMSLSTHSSVQAGWKLPPNLSPPTIEYNDG 2520
QY 2521 RPAKRHDIAARSHSESRLPPINRSGTWKREHSHSSSLPRVSTWRTGSSSILSASSES 2580
Db 2521 RPAKRHDIAARSHSESRLPPINRSGTWKREHSHSSSLPRVSTWRTGSSSILSASSES 2580
QY 2581 SEKAKSEDEKHVNSISGTSQKSNQVSAKGTWRKIKENEFSPNTSQTQVSSGATNGAES 2640
Db 2581 SEKAKSEDEKHVNSISGTSQKSNQVSAKGTWRKIKENEFSPNTSQTQVSSGATNGAES 2640
QY 2641 KTLIYQMAFAVAKTDEWVRIEDCPINPRSGRSPGTGNTPPVIDSVSEKANENIKDSXDN 2700
Db 2641 KTLIYQMAFAVAKTDEWVRIEDCPINPRSGRSPGTGNTPPVIDSVSEKANENIKDSXDN 2700
QY 2701 QAKQNVGSGVPMRTVGLNRLNSFTQVDAPOKGTETKPGQNPVPVSETNESSIVERT 2760
Db 2701 QAKQNVGSGVPMRTVGLNRLNSFTQVDAPOKGTETKPGQNPVPVSETNESSIVERT 2760
QY 2761 PFSSSSSSXHSFSGTVAARVTPFNPNPSRKSSADSTSARPSQIPTPVNNNTKRDGKT 2820
Db 2761 PFSSSSSSXHSFSGTVAARVTPFNPNPSRKSSADSTSARPSQIPTPVNNNTKRDGKT 2820
QY 2821 DSTESGTSQPKRHSGSYLVTGV 2843
Db 2821 DSTESGTSQPKRHSGSYLVTGV 2843

RESULT 3

AAB23011
ID AAB23011 standard; protein; 2843 AA.
AC AAB23011;
XX
DT 16-JAN-2001 (first entry)
XX
DE Human APC protein (splice variant 1).
XX

KW APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;
KW familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;
KW sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;
KW bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;
KW tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;
KW genetic predisposition; drug screening; Dp2.5; splice variant.
XX

OS Homo sapiens.

XX US6114124-A.
XX
PD 05-SEP-2000.
XX
PF 25-MAY-1995; 95US-00450582.
XX
PR 16-JAN-1991; 91GB-000000962.
PR 16-JAN-1991; 91GB-000000963.
PR 16-JAN-1991; 91GB-000000974.
PR 16-JAN-1991; 91GB-000000975.
PR 08-AUG-1991; 91US-00741940.
PR 12-AUG-1994; 94US-00289548.
XX

PA (ICIL) IMPERIAL CHEM IND PLC.
PA (UYGO) UNIV JOHNS HOPKINS.
PA (UTAH) UNIV UTAH.
PA (CANC-) CANCER INST.
XX
PI Carlsson M, Groden J, Joslyn G, Kinzler K, Markham AF, Anand R;
PI Albertsen H, White RL, Thliveris A, Nakamura Y, Vogelstein B;
PI Hedge PJ;
XX
DR WPI; 2000-565003/52.
DR N-PSDB; AAA93449.
XX

PT Detecting Adenomatous Polyposis Coli (APC) protein in a sample for
PT diagnosing cancers, involves contacting the sample with antibodies that
PT specifically bind to APC protein and detecting the complex formed.
XX
PS Claim 1; Fig 3A-C; 125pp; English.
XX
CC The invention relates to a novel method for detecting Adenomatous
CC Polyposis Coli (APC) protein in a sample. The method involves contacting
CC the sample with antibodies which specifically binds to the 2843 amino
CC acid form of the human APC protein, or to a mutant APC protein, and
CC detecting an APC-antibody complex. Mutations in the APC gene play a role
CC in tumorigenesis, indicating that it is a tumour suppressor gene. It is
CC located on chromosome 5q21, which corresponds to the FAP (familial
CC adenomatous polyposis) locus. FAP is an autosomal dominant inherited
CC disease in which affected individuals develop hundreds to thousands of
CC adenomatous polyps in the colon and rectum, some of which progress to
CC malignancy. The FAP locus is often found to be deleted in sporadic (i.e.,
CC non-familial) adenomas and carcinomas, and chromosome 5q deletions have
CC also been observed in tumours of the lung, breast, colon, rectum,
CC bladder, liver, sarcoma, stomach, and prostate, and in leukaemias and
CC lymphomas. Although the FAP locus contains several other genes such as
CC FER, TBL1, TB2, and MCC, it is thought that mutations in the APC gene play
CC a key role in the development of FAP and sporadic tumours. The method is
CC useful for detecting APC protein and its mutant forms in foetal tissue,
CC placental tissue, amniotic fluid, blood, serum or a tumour sample. The
CC method is useful for diagnosing or prognosing neoplastic tissue, for
CC detecting a genetic predisposition to cancer, for detecting germline and
CC somatic alteration of wild-type APC genes, and for testing therapeutic
CC agents for the ability to suppress tumours. The present sequence
CC represents a 2843 amino acid splice variant of the human APC protein.
CC This variant is more abundant than the 2742 amino acid variant (AAB23012)
XX
SQ Sequence 2843 AA;

Query Match 100.0%; Score 14566; DB 3; Length 2843;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAASVYDQLLKQVEALKWENSLRQLEDNSNHLTKLETSASNKVKVLKQGSIEDEAM 60
Db 1 MAASVYDQLLKQVEALKWENSLRQLEDNSNHLTKLETSASNKVKVLKQGSIEDEAM 60
Qy 61 ASSQIDILLRLKELNLDSSNFFGVKLRSMKRSYSGREGSVSSRSGEGSPVPMGSFPR 120
Db 61 ASSQIDILLRLKELNLDSSNFFGVKLRSMKRSYSGREGSVSSRSGEGSPVPMGSFPR 120
Qy 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKQWYVLAQLNLTKRIDSLPTENFSL 180
Db 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKQWYVLAQLNLTKRIDSLPTENFSL 180
Qy 181 QTDWTRRQLEYEARQIRVAMEEQGTCCQDMKEKAQRRIARIQIEXDILRIQLLOSQAT 240
Db 181 QTDWTRRQLEYEARQIRVAMEEQGTCCQDMKEKAQRRIARIQIEXDILRIQLLOSQAT 240
Qy 241 EAERSSONKHETGSHDAERONEGGVGEINWATSGNGQSTTMDHETASVLSSTSTHA 300
Db 241 EAERSSONKHETGSHDAERONEGGVGEINWATSGNGQSTTMDHETASVLSSTSTHA 300
Qy 301 PRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSSSQDSCISNRSGCPLLIQLL 360
Db 301 PRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSSSQDSCISNRSGCPLLIQLL 360
Qy 361 HGNDKDSVLLGNSRSGKEAPARASAAHNIHSQPDCKRGRRIIRVHLLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNSRSGKEAPARASAAHNIHSQPDCKRGRRIIRVHLLEQIRAYCETC 420
Qy 421 WEQEAHEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGCAIAELQ 480
Db 421 WEQEAHEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGCAIAELQ 480
Qy 481 VDCMYGLTNDHYSITLRRYAGMALNLTFGDVANKATLCSMKGCNRALVAQKSESEDL 540
Db 481 VDCMYGLTNDHYSITLRRYAGMALNLTFGDVANKATLCSMKGCNRALVAQKSESEDL 540

541 QOVIASVLRLSWRADVNSKTLREYGSVKALMECALEVKESTLKSVLALWNLSAHCT 600
 541 QOVIASVLRLSWRADVNSKTLREYGSVKALMECALEVKESTLKSVLALWNLSAHCT 600
 601 ENKADI CAVDGALAFVLGTLTYRSQNTLAIIESGGGILRNVSLLIATNEDHQILRENN 660
 601 ENKADI CAVDGALAFVLGTLTYRSQNTLAIIESGGGILRNVSLLIATNEDHQILRENN 660
 661 CLQTLQHLKSHSLTIVSNACGLTWNLSARNPKDQALWDMGAVSMLKXLIHSKXMIAM 720
 661 CLQTLQHLKSHSLTIVSNACGLTWNLSARNPKDQALWDMGAVSMLKXLIHSKXMIAM 720
 721 GSAALANLWANRPAYKDANIMSPGSSLPSSHVRKQKALEABLDQHLSETFDNIDNLS 780
 721 GSAALANLWANRPAYKDANIMSPGSSLPSSHVRKQKALEABLDQHLSETFDNIDNLS 780
 781 PKASHRSKQHKOSLYGDIYVDFNRHDDNPSDNFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
 781 PKASHRSKQHKOSLYGDIYVDFNRHDDNPSDNFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
 841 LDSRSSEKDRSLRERIGIGLNYHPATENPGTSSKRGLOISTTAAQIAKVMEEVSAIHTS 900
 841 LDSRSSEKDRSLRERIGIGLNYHPATENPGTSSKRGLOISTTAAQIAKVMEEVSAIHTS 900
 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHNTYNTKSENSNRTCSMPYAKLEYKRS 960
 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHNTYNTKSENSNRTCSMPYAKLEYKRS 960
 961 NDSLSNVSSSDGCGKQOMKPSIESYSEDDDESFCYGOYPADLAHKIHSANHMDNDGE 1020
 961 NDSLSNVSSSDGCGKQOMKPSIESYSEDDDESFCYGOYPADLAHKIHSANHMDNDGE 1020
 1021 LDTPINYSKYSLYDBQLNGRQSPQNERWARPKHIIIEDEIKQSEQRQSNQSTTYPVYTE 1080
 1021 LDTPINYSKYSLYDBQLNGRQSPQNERWARPKHIIIEDEIKQSEQRQSNQSTTYPVYTE 1080
 1081 STDDKHLKFQHFQOQCVFPYRSRGANGSETNVRGSHGINQVSGSLCQEDDYEDDKP 1140
 1081 STDDKHLKFQHFQOQCVFPYRSRGANGSETNVRGSHGINQVSGSLCQEDDYEDDKP 1140
 1141 TNYSERYSSEEEHBEERPTNYSIKYNEKRHVDPIDYSLKYATDIPSSQKQSPFSKS 1200
 1141 TNYSERYSSEEEHBEERPTNYSIKYNEKRHVDPIDYSLKYATDIPSSQKQSPFSKS 1200
 1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINOETIQ 1260
 1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINOETIQ 1260
 1261 TYCVEDTPICFRSCSLSSLSAASDEIGCQNTQOADSANTLOIAETKEKIGTRSAEDPV 1320
 1261 TYCVEDTPICFRSCSLSSLSAASDEIGCQNTQOADSANTLOIAETKEKIGTRSAEDPV 1320
 1321 SEVPAVSQHPKTSRLOGSLSSSARHKAVERSSGAKSPSKGAOTPKSPPEHYVOET 1380
 1321 SEVPAVSQHPKTSRLOGSLSSSARHKAVERSSGAKSPSKGAOTPKSPPEHYVOET 1380
 1381 PLMFSRCTSVSSLSDFSRSTASSVQSEPCSGMVSGIISPDLDPSPGQWMPFSRSKTPP 1440
 1381 PLMFSRCTSVSSLSDFSRSTASSVQSEPCSGMVSGIISPDLDPSPGQWMPFSRSKTPP 1440
 1441 PPPQTAQKREVPKNAKTAPEKRESGKQAAVNAVQVLPDADTLHLFATESPDGF 1500
 1441 PPPQTAQKREVPKNAKTAPEKRESGKQAAVNAVQVLPDADTLHLFATESPDGF 1500
 1501 SCSSLSALSJDEEPIQKQVLRIMPVQENDNGNETESEOPEKSENQKEAEKTIIDSE 1560
 1501 SCSSLSALSJDEEPIQKQVLRIMPVQENDNGNETESEOPEKSENQKEAEKTIIDSE 1560
 1561 KDLLDDSDDDDEIILEECIIISAMPTKSRKAKQAQATASKLPPPVARKPSQLPYKLLPS 1620
 1561 KDLLDDSDDDDEIILEECIIISAMPTKSRKAKQAQATASKLPPPVARKPSQLPYKLLPS 1620

QY 1621 QNRLOPQKHVSFTPGDDMPRVYCVGEGTPIINFSTATISLDLTIESPPNELAAGGVRGGAQ 1680
 DB 1621 QNRLOPQKHVSFTPGDDMPRVYCVGEGTPIINFSTATISLDLTIESPPNELAAGGVRGGAQ 1680
 QY 1681 SGEPEKRDITPTGRSDTDEAQQGKTSSVTIPELDNKAEBGDILAEICINSAMPKGSHPK 1740
 DB 1681 SGEPEKRDITPTGRSDTDEAQQGKTSSVTIPELDNKAEBGDILAEICINSAMPKGSHPK 1740
 QY 1741 FRVKIMDOVQOASASSAPNKNQLDGKKKPTSPVKPIPONTTEYTRVRKNADSKNNLN 1800
 DB 1741 FRVKIMDOVQOASASSAPNKNQLDGKKKPTSPVKPIPONTTEYTRVRKNADSKNNLN 1800
 QY 1801 AERVFSDNKSQKQNLKNNKDFNDKLPNNEDVRGSAFDSPHHTPIEGTYPYCFSRND 1860
 DB 1801 AERVFSDNKSQKQNLKNNKDFNDKLPNNEDVRGSAFDSPHHTPIEGTYPYCFSRND 1860
 QY 1861 SLSLDFDDDDVDLSREKAEIRKAKENKESAEKVTSHTELTSNQOSANKTOAIKQPINR 1920
 DB 1861 SLSLDFDDDDVDLSREKAEIRKAKENKESAEKVTSHTELTSNQOSANKTOAIKQPINR 1920
 QY 1921 QCPKPILOKQSTFPQSSKDIPIRGAAATDEKLQNPATIENTPVCFSHNSLSLSLSDIDQENN 1980
 DB 1921 QCPKPILOKQSTFPQSSKDIPIRGAAATDEKLQNPATIENTPVCFSHNSLSLSLSDIDQENN 1980
 QY 1981 NKENEPKETEPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNSLSLSLSDIDQENN 2040
 DB 1981 NKENEPKETEPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNSLSLSLSDIDQENN 2040
 QY 2041 ECIISAMPKPKKRLKGDNEKHSRNMGGTGLGDLTLDLKDQORPDSEHGLSPDSNED 2100
 DB 2041 ECIISAMPKPKKRLKGDNEKHSRNMGGTGLGDLTLDLKDQORPDSEHGLSPDSNED 2100
 QY 2101 WKATOEGANSIVSLHQAAAAACLSRQASDSDSILSKSGISLGSPPHLPDQERKPT 2160
 DB 2101 WKATOEGANSIVSLHQAAAAACLSRQASDSDSILSKSGISLGSPPHLPDQERKPT 2160
 QY 2161 SNKGPRILKPEKSTLETETKIESKGIKGGKVKYKSLITGKVRNSSEISQMKQPIQAN 2220
 DB 2161 SNKGPRILKPEKSTLETETKIESKGIKGGKVKYKSLITGKVRNSSEISQMKQPIQAN 2220
 QY 2221 MFSISRGTMTHIPGVNRSSSTSPVSKGKPLKTPAKSPSEGTATTPSPRGAKPSVKS 2280
 DB 2221 MFSISRGTMTHIPGVNRSSSTSPVSKGKPLKTPAKSPSEGTATTPSPRGAKPSVKS 2280
 QY 2281 ELSPVARQTSQIGSSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPRNGISPPN 2340
 DB 2281 ELSPVARQTSQIGSSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPRNGISPPN 2340
 QY 2341 KLSQLPRTSSPSTASTKSSGSKYTSPPQMSQOQNLTKOTGLSKXASSIPRSESASKG 2400
 DB 2341 KLSQLPRTSSPSTASTKSSGSKYTSPPQMSQOQNLTKOTGLSKXASSIPRSESASKG 2400
 QY 2401 LQNMNGANGANKVLSRYSTKSSGESDRSERPVLVRQSTFTKEAPSPTRKLEESA 2460
 DB 2401 LQNMNGANGANKVLSRYSTKSSGESDRSERPVLVRQSTFTKEAPSPTRKLEESA 2460
 QY 2461 SPESLSPSRSPASPTRSQOATPVLSPSLPDMSLSTHSSVQAGWKLPNLSPIETENDG 2520
 DB 2461 SPESLSPSRSPASPTRSQOATPVLSPSLPDMSLSTHSSVQAGWKLPNLSPIETENDG 2520
 QY 2521 RPAKHDIARSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
 DB 2521 RPAKHDIARSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
 QY 2581 SEKAKSEDEKHVNSISGTSKQKNOVSAGTWKTKENEFPTNSTQTSVSGATNGAES 2640
 DB 2581 SEKAKSEDEKHVNSISGTSKQKNOVSAGTWKTKENEFPTNSTQTSVSGATNGAES 2640
 QY 2641 KTLIYQMAPAVSKTEDVWVRIBDCPFINNPRSGRSPGTNTPPVIDSVSEKANPNIKSDKN 2700
 DB 2641 KTLIYQMAPAVSKTEDVWVRIBDCPFINNPRSGRSPGTNTPPVIDSVSEKANPNIKSDKN 2700
 QY 2701 QAKQNVGSGVPMRTVGLNRLNSFIQVADPDQKTEIKFGQNNPVPVSEINNESSIVERT 2760

Db 2701 CAQNVGSGVPMRTVGVLEBINSIQVDADQKGTETKPGQNPVPVSETNESSIVERT 2760
 QY 2761 PFSSSSSKHSPSGTVAARTVTFPNYNSPRKSSADSTSARPSQIPTPPVNNTKRDSKT 2820
 Db 2761 PFSSSSSKHSPSGTVAARTVTFPNYNSPRKSSADSTSARPSQIPTPPVNNTKRDSKT 2820
 QY 2821 DSTESSGTQSPKRHSGSYLVTSV 2843
 Db 2821 DSTESSGTQSPKRHSGSYLVTSV 2843

RESULT 4
 ABG71105
 ID ABG71105 standard; protein; 2843 AA.
 AC ABG71105;
 DT 08-JAN-2003 (first entry)
 DE Human adenomatous polyposis coli (APC) protein.
 XX Human; adenomatous polyposis coli; APC; transgenic animal; colon cancer;
 KW developmental abnormality; anticancer drug; beta-catenin;
 KW Wnt/Wg signalling pathway.
 XX Homo sapiens.
 OS Homo sapiens.
 FN EPI243646-A2.
 PD 25-SEP-2002.
 XX 07-MAR-2002; 2002EP-00290571.
 XX 19-MAR-2001; 2001US-0276483P.
 XX (COUL) COUNCIL SCI & IND RES.
 PI Bhandari P, Shashidhara LS;
 DR WPI; 2002-742695/81.

New transgenic Drosophila containing the human colon cancer gene
 Adenomatous Polyposis Coli is useful as an assay model to screen for new
 drugs, particularly against colon cancer.
 Claim 1; Page 19-20; 52pp; English.

The invention describes a transgenic Drosophila whose genome comprises
 the full-length human colon cancer gene Adenomatous Polyposis Coli (APC)
 allowing regulated mis-expression of the APC gene resulting in
 developmental abnormalities. The transgenic flies are used to screen and
 validate efficacy of anticancer drugs, to identify new target proteins
 interacting with beta-catenin, genes which interact with human APC, to
 study the biochemical function of human APC and to identify additional
 components of the Drosophila Wnt/Wg signalling pathway. In particular the
 flies are used to screen potential drugs against colon cancer. This is
 the amino acid sequence of the human adenomatous polyposis coli (APC)
 protein DNA encoding which is incorporated in the transgenic flies
 Sequence 2843 AA;

Query Match 100.0%; Score 14566; DB 5; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAASYDQLLKQVEALKMNSNLRQLEDNSNHLTKLETEASNKKEVLKQCGSIEDEAM 60
 Db 1 MAAASYDQLLKQVEALKMNSNLRQLEDNSNHLTKLETEASNKKEVLKQCGSIEDEAM 60
 QY 61 ASSGQIDLLERLKEINLDSNFPFGVKLRKSKMSLRSGREGSVSRSGECSFVPMGSPFR 120
 Db 61 ASSGQIDLLERLKEINLDSNFPFGVKLRKSKMSLRSGREGSVSRSGECSFVPMGSPFR 120

QY 121 RGFVNGSRSTGYLELEKERSLLADLDKEEKEKDWYQAQLNLTAKRIDSLPTNFSL 180
 Db 121 RGFVNGSRSTGYLELEKERSLLADLDKEEKEKDWYQAQLNLTAKRIDSLPTNFSL 180
 QY 181 QTMTRRQLEYEARQIRVAMEEQGTGCDMEKRAQRIARIQOIEKDIIRIROLLOSOAT 240
 Db 181 QTMTRRQLEYEARQIRVAMEEQGTGCDMEKRAQRIARIQOIEKDIIRIROLLOSOAT 240
 QY 241 EAERSSONKHETGSHDAERQNEQGVGEINMATSNGQSGSTTMDHETASVLSSSSTHSA 300
 Db 241 EAERSSONKHETGSHDAERQNEQGVGEINMATSNGQSGSTTMDHETASVLSSSSTHSA 300
 QY 301 PRELTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSSQSCISMRSGCPLLIQLL 360
 Db 301 PRELTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSSQSCISMRSGCPLLIQLL 360
 QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIISQPDCKRGRREIRVHLLEQIRAYCETC 420
 Db 361 HGNDKDSVLLGNSRGSKEARASAAALHNIISQPDCKRGRREIRVHLLEQIRAYCETC 420
 QY 421 WEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 480
 Db 421 WEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 480
 QY 481 VDCENYGLTNDHYSITLRYAGMALTNLTFGVDVANKATILCSMKGCMRALVAQIKSSEDL 540
 Db 481 VDCENYGLTNDHYSITLRYAGMALTNLTFGVDVANKATILCSMKGCMRALVAQIKSSEDL 540
 QY 541 QQVIASVLRNLSWRAADVNSKKTIREVGSVKALMECALEVKESTLKSLSALWNLSAHCT 600
 Db 541 QQVIASVLRNLSWRAADVNSKKTIREVGSVKALMECALEVKESTLKSLSALWNLSAHCT 600
 QY 601 ENKADICAVDGAFLVGLTYRSQNTLAIIBSGGILRNVSLLIATNEDHROIIRENN 660
 Db 601 ENKADICAVDGAFLVGLTYRSQNTLAIIBSGGILRNVSLLIATNEDHROIIRENN 660
 QY 661 CLQTLQHLKSHSLTIVSNACGLTNLSARNPKDQALWDMGVSMKLIHSHKHVMAM 720
 Db 661 CLQTLQHLKSHSLTIVSNACGLTNLSARNPKDQALWDMGVSMKLIHSHKHVMAM 720
 QY 721 GSAALRLNLMANRPKAKYDANIMSPGSLPSLHVRKQKALEABDLAQHLSSETFDNDLS 780
 Db 721 GSAALRLNLMANRPKAKYDANIMSPGSLPSLHVRKQKALEABDLAQHLSSETFDNDLS 780
 QY 781 PKASHRSKQHKOSLYGCVYVDNTHDDNSDNFNTGNMTVLSPLYNTTVLPSSSSRGS 840
 Db 781 PKASHRSKQHKOSLYGCVYVDNTHDDNSDNFNTGNMTVLSPLYNTTVLPSSSSRGS 840
 QY 841 LDSRSSEKORSLEERERIGLGNYPATENPGTSSKGLQITTTAAQIAKWEVSAIHTS 900
 Db 841 LDSRSSEKORSLEERERIGLGNYPATENPGTSSKGLQITTTAAQIAKWEVSAIHTS 900
 QY 901 QEDRSSGTTTELHCVTDERNALRSSAAHSTNTYFTKSENSNRCTCSMPYAKLEYKRSS 960
 Db 901 QEDRSSGTTTELHCVTDERNALRSSAAHSTNTYFTKSENSNRCTCSMPYAKLEYKRSS 960
 QY 961 NDSLSVSSSDGYGKKGQMKPSTESYSEDESKFCYGOVPADLAHKIHSANHMDNDGE 1020
 Db 961 NDSLSVSSSDGYGKKGQMKPSTESYSEDESKFCYGOVPADLAHKIHSANHMDNDGE 1020
 QY 1021 LDTPINYSKYDEQLNSGRQSPSONERWAPKHIIEDEIKOSEQRQSNRQSTTYPVYTE 1080
 Db 1021 LDTPINYSKYDEQLNSGRQSPSONERWAPKHIIEDEIKOSEQRQSNRQSTTYPVYTE 1080
 QY 1081 STDDKHLKQPHGQCEVSPYBSRGANGSETNRVGSNHGINONVSQSCICEQDDYEDDKP 1140
 Db 1081 STDDKHLKQPHGQCEVSPYBSRGANGSETNRVGSNHGINONVSQSCICEQDDYEDDKP 1140
 QY 1141 TNYSERYSSEEQHEEERPTNYSIKYNEEKRVHDQPIDYSLKYATDIPSSQKQSFSSKS 1200
 Db 1141 TNYSERYSSEEQHEEERPTNYSIKYNEEKRVHDQPIDYSLKYATDIPSSQKQSFSSKS 1200

QY 1201 SSGSSKTEHSSSSSENTSTPSSNAKRONQLHPSSAQSRSGQPKAATCKVKSINQETIQ 1260
 Db 1201 SSGSSKTEHSSSSSENTSTPSSNAKRONQLHPSSAQSRSGQPKAATCKVKSINQETIQ 1260
 QY 1261 TYCVEDTPICFSRCSSLSLSAEDEIGCQNTQCEADSAANTLQIABIKIGTRSAEDPV 1320
 Db 1261 TYCVEDTPICFSRCSSLSLSAEDEIGCQNTQCEADSAANTLQIABIKIGTRSAEDPV 1320
 QY 1321 SEVPANVCHPRTKSSRLQGSLSLSAEARHKAFFSSGAKSPSSGACTPKSPPEHYVQET 1380
 Db 1321 SEVPANVCHPRTKSSRLQGSLSLSAEARHKAFFSSGAKSPSSGACTPKSPPEHYVQET 1380
 QY 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPGQTMPSPRSKTPP 1440
 Db 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPGQTMPSPRSKTPP 1440
 QY 1441 PPQTAQTKREVKNKAPTAEKESGPKQAANVAARVQVLPDADTLHFAESTPDGF 1500
 Db 1441 PPQTAQTKREVKNKAPTAEKESGPKQAANVAARVQVLPDADTLHFAESTPDGF 1500
 QY 1501 SCSSLSALSILDEPFIQKVELRIMPVOENDNGNETESQPKESNENQEKAEKTIIDSE 1560
 Db 1501 SCSSLSALSILDEPFIQKVELRIMPVOENDNGNETESQPKESNENQEKAEKTIIDSE 1560
 QY 1561 KDLLDDSDDDDIIELEECIIISAMPTKSSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
 Db 1561 KDLLDDSDDDDIIELEECIIISAMPTKSSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
 QY 1621 QNRLOPKXHSVFTPGDDMPVYCVGFTPIFNSFATSLSDLTIESPNELAAAGEVRGGQA 1680
 Db 1621 QNRLOPKXHSVFTPGDDMPVYCVGFTPIFNSFATSLSDLTIESPNELAAAGEVRGGQA 1680
 QY 1681 SGFEKEDTITPTEGRSTDEAGGKTSSVTIPELDNKAEEGDILABCINSAMPKGSHPK 1740
 Db 1681 SGFEKEDTITPTEGRSTDEAGGKTSSVTIPELDNKAEEGDILABCINSAMPKGSHPK 1740
 QY 1741 FRVKKINDVQOASASSAPKNQLOGKKKPTSPVKPIPONTEYTRVRKNADSKNNLN 1800
 Db 1741 FRVKKINDVQOASASSAPKNQLOGKKKPTSPVKPIPONTEYTRVRKNADSKNNLN 1800
 QY 1801 AERFSDNKDKKQNLKNNKDFNDKLPNNEDVRGSAFDPSPHHYTPIEGTPYCFSRND 1860
 Db 1801 AERFSDNKDKKQNLKNNKDFNDKLPNNEDVRGSAFDPSPHHYTPIEGTPYCFSRND 1860
 QY 1861 SLSLDDDDDDVLSREKAEKLRKAKENKESAKVTSHTELTSNQOANKTQAIKQPINR 1920
 Db 1861 SLSLDDDDDDVLSREKAEKLRKAKENKESAKVTSHTELTSNQOANKTQAIKQPINR 1920
 QY 1921 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
 Db 1921 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
 QY 1981 NKNEPIKETEPDPSQSEPKQASGAPKSFHVEDTPVCFNSNSLSLSLSDIDSEDDLLQ 2040
 Db 1981 NKNEPIKETEPDPSQSEPKQASGAPKSFHVEDTPVCFNSNSLSLSLSDIDSEDDLLQ 2040
 QY 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGILGEDITLDDKIQRPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGILGEDITLDDKIQRPDSEHGLSPDSENF 2100
 QY 2101 WKAIQEGANSIVSLHQAACCLSRQASSDSDSILSLKSGISLGSFPFHLTPDQEKPT 2160
 Db 2101 WKAIQEGANSIVSLHQAACCLSRQASSDSDSILSLKSGISLGSFPFHLTPDQEKPT 2160
 QY 2161 SNKGPRILKXGKSTLTETKIESKIGKGVKYSKLIITGKVRNSSEISGQMKQPLQAN 2220
 Db 2161 SNKGPRILKXGKSTLTETKIESKIGKGVKYSKLIITGKVRNSSEISGQMKQPLQAN 2220
 QY 2221 MPGISRGRTMIHIPGVNNSSSSTSPVSKGPPPLKTPASKSPSEGTATTSRQAKPSVKS 2280
 Db 2221 MPGISRGRTMIHIPGVNNSSSSTSPVSKGPPPLKTPASKSPSEGTATTSRQAKPSVKS 2280
 QY 2281 ELSPVARQTSIQGSSKAPSRSGSRDSTPSRPAQPLSRPIQSPGRNSISPGENGISPPN 2340

Db 2281 ELSPVARQTSIQGSSKAPSRSGSRDSTPSRPAQPLSRPIQSPGRNSISPGENGISPPN 2340
 QY 2341 KLSQIPRTSSPSTASTKSSSGSKMSYTPSGEQSQONLTQTGLSKNASSIPRSESASKG 2400
 Db 2341 KLSQIPRTSSPSTASTKSSSGSKMSYTPSGEQSQONLTQTGLSKNASSIPRSESASKG 2400
 QY 2401 LNQNNNGANKKVELSRMSTKSSSGSDRSERPVLVROSTFIKEAPSPTLRKLKLESA 2460
 Db 2401 LNQNNNGANKKVELSRMSTKSSSGSDRSERPVLVROSTFIKEAPSPTLRKLKLESA 2460
 QY 2461 SFESLSPSRSPASTRQAOQTPVLSPLPDKSLSTHSSVQAGGWRKLPPLSPTEIYNDG 2520
 Db 2461 SFESLSPSRSPASTRQAOQTPVLSPLPDKSLSTHSSVQAGGWRKLPPLSPTEIYNDG 2520
 QY 2521 RPAKRHDIAIARSHSPRLPINRSGTWKRSHSHSSLPVSTWRTGSSSSILSASSES 2580
 Db 2521 RPAKRHDIAIARSHSPRLPINRSGTWKRSHSHSSLPVSTWRTGSSSSILSASSES 2580
 QY 2581 SEKAKSDEKHVNSISGTSKQKSNQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 Db 2581 SEKAKSDEKHVNSISGTSKQKSNQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 QY 2641 KTLIYQMAPAVSKTEDVWRIEDCPINNPSGRSPTGNTTPVIDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIYQMAPAVSKTEDVWRIEDCPINNPSGRSPTGNTTPVIDSVSEKANPNIKDSKN 2700
 QY 2701 QAKQNVGSGVPMRTVGLNRLNSFIQVADAPDQKGTIKFGQNNPVPVSETNESSIVERT 2760
 Db 2701 QAKQNVGSGVPMRTVGLNRLNSFIQVADAPDQKGTIKFGQNNPVPVSETNESSIVERT 2760
 QY 2761 PFSSSSSKHSPGCTVAARVTPENYNPSRKSSADSTARSPOIPTPVANNTKKRDST 2820
 Db 2761 PFSSSSSKHSPGCTVAARVTPENYNPSRKSSADSTARSPOIPTPVANNTKKRDST 2820
 QY 2821 DSTESSGTQSPKSHSGSYLTVTSV 2843
 Db 2821 DSTESSGTQSPKSHSGSYLTVTSV 2843
 RESULT 5
 ID AAW76821 standard; protein; 2973 AA.
 AC AAW76821;
 DT 25-JAN-1999 (first entry)
 DE Human APC protein.
 KW Tcf; beta-catenin; human; drug; familial adenomatous polyposis; FAP;
 KW cancer; adenomatous polyposis coli; APC; neoplastic.
 OS Homo sapiens.
 PN WO9841631-A2.
 PD 24-SEP-1998.
 PF 20-MAR-1998; 98WO-US005506.
 PR 20-MAR-1997; 97US-00821355.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UYUT-) RIJKSUNIV UTRECHT.
 PI Barker N, Clevers H, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
 PI Vogelstein B;
 WPI; 1998-531569/45.
 PT Intron-free DNA encoding Tcf-4 protein - useful for, e.g. identifying
 drugs for treating FAP patients, or patients with increased risk of

PT developing cancer.
XX Disclosure; Page 32-37; 58pp; English.
XX This sequence represents a human APC protein which is used in a method of
CC identifying candidate drugs for use in familial adenomatous polyposis
CC (FAP) patients, or patients with increased risk of developing cancer. The
CC protein can also be used to determine the presence or absence in a cell
CC of wild type adenomatous polyposis coli (APC) gene or a downstream
CC protein in the APC transcriptional regulatory pathway. This method involves
CC introducing a Tcf-responsive reporter gene into the cell, and measuring
CC transcription of the reporter gene where a cell which supports active
CC transcription of the reporter gene does not have wild type APC or a
CC downstream protein of the APC transcriptional regulatory pathway. The
CC protein can also be used in a method of diagnosing cancer in a sample
CC suspected of being neoplastic and for treating a patient with colorectal
CC cancer or other cancer associated with FAP, comprising administering to
CC the patient a polypeptide comprising a portion of the APC sequence,
CC comprising the beta-catenin binding site
XX Sequence 2973 AA;

Query Match		100.0%;	Score 14566;	DB 2;	Length 2973;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2843;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAAAYDQLLKQVEALXVENSRLQELNLSNHLTKLETEASNMKEVLKQ-QGSIEDAM	60		
DB	1	MAAAYDQLLKQVEALXVENSRLQELNLSNHLTKLETEASNMKEVLKQ-QGSIEDAM	60		
QY	61	ASSGOIDLLERLKLNLDSNPPGVKLRKMSLRSGREGSVSRSGECSPVPMGSPPR	120		
DB	61	ASSGOIDLLERLKLNLDSNPPGVKLRKMSLRSGREGSVSRSGECSPVPMGSPPR	120		
QY	121	RGVNGSRESTGYLEBEKERSLLADLDKEEKXDWYAOQLNLTIRIDSLPTENPSL	180		
DB	121	RGVNGSRESTGYLEBEKERSLLADLDKEEKXDWYAOQLNLTIRIDSLPTENPSL	180		
QY	181	QTDMTROLEYEARQIRVAMEEQLGTCQDMKRAQRRIARIQIQEKDILRLQLQSQT	240		
DB	181	QTDMTROLEYEARQIRVAMEEQLGTCQDMKRAQRRIARIQIQEKDILRLQLQSQT	240		
QY	241	EAERSSQNHGTGSHDAERQEGQVGEINMATSNGOGSTTRMDHETASVLSSTHSA	300		
DB	241	EAERSSQNHGTGSHDAERQEGQVGEINMATSNGOGSTTRMDHETASVLSSTHSA	300		
QY	301	PRRLTSHLGTVMYVYLLSMLGTHDKDDMGRTLLIAMSSODSCTSMRQSGCLPLLQLL	360		
DB	301	PRRLTSHLGTVMYVYLLSMLGTHDKDDMGRTLLIAMSSODSCTSMRQSGCLPLLQLL	360		
QY	361	HGNDKSVLLGNRSGSKARARASALHNIITHSQDDKRGREIRVLHLEQIRAYCETC	420		
DB	361	HGNDKSVLLGNRSGSKARARASALHNIITHSQDDKRGREIRVLHLEQIRAYCETC	420		
QY	421	WEWQEAHBFQMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRRHAMELGLQIAELLQ	480		
DB	421	WEWQEAHBFQMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRRHAMELGLQIAELLQ	480		
QY	481	VDCEMYGLTNDHYITLRRYAGMALTNLTTFQDVANKATLCSMKCMRALVAQLKSEEDL	540		
DB	481	VDCEMYGLTNDHYITLRRYAGMALTNLTTFQDVANKATLCSMKCMRALVAQLKSEEDL	540		
QY	541	QQVIASVLRLNLSRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALMNLSAHCT	600		
DB	541	QQVIASVLRLNLSRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALMNLSAHCT	600		
QY	601	ENKADICAVDGALAFVGLTYRQTNLTALIESGGGILRVNSSLIATNEDHRIQLENN	660		
DB	601	ENKADICAVDGALAFVGLTYRQTNLTALIESGGGILRVNSSLIATNEDHRIQLENN	660		
QY	661	CLQTLLOHLKSHSLTIVSNACGTLWNLSARNPKDQBALMDWGVSMKLNLIHSHKGMAM	720		
DB	661	CLQTLLOHLKSHSLTIVSNACGTLWNLSARNPKDQBALMDWGVSMKLNLIHSHKGMAM	720		

QY	721	GSAALRNLMANRPAYKNDANINSPGSSLPSPSHVRKQKALAEALDAQHLSETFNIDNLS	780
DB	721	GSAALRNLMANRPAYKNDANINSPGSSLPSPSHVRKQKALAEALDAQHLSETFNIDNLS	780
QY	781	PKASHRSKQKHQSLYGDYVFTNRHDDNDSNPTGNTVLSYLTATTVLPSSSSSRGS	840
DB	781	PKASHRSKQKHQSLYGDYVFTNRHDDNDSNPTGNTVLSYLTATTVLPSSSSSRGS	840
QY	841	LDSSRSKDSRLERERIGLGNYPATENPQTSKRGIGQISTTAAQIAKMWEEVSAIHTS	900
DB	841	LDSSRSKDSRLERERIGLGNYPATENPQTSKRGIGQISTTAAQIAKMWEEVSAIHTS	900
QY	901	QEDSSSGTTELHCVTDERNALRRSSAAHNTNTYNTKSENSNRTCSMPYAKLEYKRSS	960
DB	901	QEDSSSGTTELHCVTDERNALRRSSAAHNTNTYNTKSENSNRTCSMPYAKLEYKRSS	960
QY	961	NDSLSNVSSSDGYGKRGOMKPSIESYEDDESFCYSQYQPADLAHKIHSANHMDNDGE	1020
DB	961	NDSLSNVSSSDGYGKRGOMKPSIESYEDDESFCYSQYQPADLAHKIHSANHMDNDGE	1020
QY	1021	LDTPINYSKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE	1080
DB	1021	LDTPINYSKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE	1080
QY	1081	STDDKHLKFPQHFQCCVSPYSRGANGSETNRVGSNHGINQVNSQSLCOEDDYEDDKP	1140
DB	1081	STDDKHLKFPQHFQCCVSPYSRGANGSETNRVGSNHGINQVNSQSLCOEDDYEDDKP	1140
QY	1141	TNYSERYSEEQHEEERPTNYSIKYNEEKRVDPQIDYSLKYATDIPSSQKQSFSPSKS	1200
DB	1141	TNYSERYSEEQHEEERPTNYSIKYNEEKRVDPQIDYSLKYATDIPSSQKQSFSPSKS	1200
QY	1201	SSGQSSKTEHMSSESTSTPSSNAKQNLHPSSAQSRGQPKAATCKVSSINQETIQ	1260
DB	1201	SSGQSSKTEHMSSESTSTPSSNAKQNLHPSSAQSRGQPKAATCKVSSINQETIQ	1260
QY	1261	TYCVEDTPICFSCSSLSLSAEDIGCNOTQEADSNLTQIAEIKKIGTRSAEDPV	1320
DB	1261	TYCVEDTPICFSCSSLSLSAEDIGCNOTQEADSNLTQIAEIKKIGTRSAEDPV	1320
QY	1321	SEVPAYSQHPRTKSKRLQSSLSSESARHKAFFSSGAKSPSKSGAQTPKSPHYVQET	1380
DB	1321	SEVPAYSQHPRTKSKRLQSSLSSESARHKAFFSSGAKSPSKSGAQTPKSPHYVQET	1380
QY	1381	PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGQTMPPSRSKTTP	1440
DB	1381	PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGQTMPPSRSKTTP	1440
QY	1441	PPQTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLHLHFATESPDGF	1500
DB	1441	PPQTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLHLHFATESPDGF	1500
QY	1501	SCSSLSLSALSLDEPFTOKDVELRIMPVQENNGNETSEQPKESNENOEKAEKTIIDSE	1560
DB	1501	SCSSLSLSALSLDEPFTOKDVELRIMPVQENNGNETSEQPKESNENOEKAEKTIIDSE	1560
QY	1561	KOLLDDSDDDIIELEECIIISAMPTKSSRKAKKPAQATASKLPPPVARKPSQLPVYKLLPS	1620
DB	1561	KOLLDDSDDDIIELEECIIISAMPTKSSRKAKKPAQATASKLPPPVARKPSQLPVYKLLPS	1620
QY	1621	QNRLOPKHVSFTPGDDMPRVYCVEGTPIINFSTATSLDITTESPPNELAAGEVGGGAQ	1680
DB	1621	QNRLOPKHVSFTPGDDMPRVYCVEGTPIINFSTATSLDITTESPPNELAAGEVGGGAQ	1680
QY	1681	SGEFEXKRTDIPTEGRSTDEAOGKTSVVITPELDDNKAEEGDILAECTINSAMPKGSHPK	1740
DB	1681	SGEFEXKRTDIPTEGRSTDEAOGKTSVVITPELDDNKAEEGDILAECTINSAMPKGSHPK	1740
QY	1741	FRVKKIMDVOQASASSAPNKNQLDGKKKCTSPVKPIQNTYRTRVRKNADSKNNLN	1800
DB	1741	FRVKKIMDVOQASASSAPNKNQLDGKKKCTSPVKPIQNTYRTRVRKNADSKNNLN	1800

1801 AERFSDNKKSKQNLKNNKDFNDKLPNNEDVRGSAFSDSPHYHTPIEGTPYCFSRND 1860
1801 AERFSDNKKSKQNLKNNKDFNDKLPNNEDVRGSAFSDSPHYHTPIEGTPYCFSRND 1860
1861 SLSSLDFFDDDDVDLSREKALRKAKENKESAKVTSHLTSTNQOSANKTQAIKQPINR 1920
1861 SLSSLDFFDDDDVDLSREKALRKAKENKESAKVTSHLTSTNQOSANKTQAIKQPINR 1920
1921 GQPKILOKQSTPPOQSKDIPORGAATDEKLNQFAIENTPVCFSHNSLSLSDIDQENN 1980
1921 GQPKILOKQSTPPOQSKDIPORGAATDEKLNQFAIENTPVCFSHNSLSLSDIDQENN 1980
1981 NKNEPIKETEPPDQSGEPKPOAGYAPKSFHVEDTPVCFERNSSLSLSDSDLDLQ 2040
1981 NKNEPIKETEPPDQSGEPKPOAGYAPKSFHVEDTPVCFERNSSLSLSDSDLDLQ 2040
2041 ECISAMPKPKPSRLKGDNEKHSPRNMGILGEDLTDLKDIQRPDSEHGLSPDSENF 2100
2041 ECISAMPKPKPSRLKGDNEKHSPRNMGILGEDLTDLKDIQRPDSEHGLSPDSENF 2100
2101 WKAIQEGANSIVSLHQAACAAACLSRQAGSDSDSILSLXSGISLGGSPFHLTPDQEKPT 2160
2101 WKAIQEGANSIVSLHQAACAAACLSRQAGSDSDSILSLXSGISLGGSPFHLTPDQEKPT 2160
2161 SNKGPRILKPEKSTLTETKIESGKIGKGVKYSLLITGKVRNSEISGQWKQLOAN 2220
2161 SNKGPRILKPEKSTLTETKIESGKIGKGVKYSLLITGKVRNSEISGQWKQLOAN 2220
2221 MPSISRGRTMIHIPGVNNSSTSPVSKGPPKLTTPASKSPSEGQTATTPRGAKEPVKS 2280
2221 MPSISRGRTMIHIPGVNNSSTSPVSKGPPKLTTPASKSPSEGQTATTPRGAKEPVKS 2280
2281 ELSFVARQTSQIGSSKAPRSGSRDTPSRPAQPLSRPIQSPGRNSISPGNGISPPN 2340
2281 ELSFVARQTSQIGSSKAPRSGSRDTPSRPAQPLSRPIQSPGRNSISPGNGISPPN 2340
2341 KLSQLPRTSPSTASTSSGSGKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASKG 2400
2341 KLSQLPRTSPSTASTSSGSGKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASKG 2400
2401 LNQMNGANGANKVELSRMSTKSSGESDRSRPVLVROSTTIKEAPSTLRKLEESA 2460
2401 LNQMNGANGANKVELSRMSTKSSGESDRSRPVLVROSTTIKEAPSTLRKLEESA 2460
2461 SFESLSPSRPASPTRSQOATPVLSLPMDSLSTHSSVQAGWRKLPPLNLSPTIEYNDG 2520
2461 SFESLSPSRPASPTRSQOATPVLSLPMDSLSTHSSVQAGWRKLPPLNLSPTIEYNDG 2520
2521 RPAKRHDIAHSHSPSRPLNIRSGTWKREHSHSSLPVSVTWRRTPGSSSSILSASSES 2580
2521 RPAKRHDIAHSHSPSRPLNIRSGTWKREHSHSSLPVSVTWRRTPGSSSSILSASSES 2580
2581 SEKAKSEDEKVNISIGTKQSKENQVAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
2581 SEKAKSEDEKVNISIGTKQSKENQVAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGTPVSDSVSEKAPNIIKSKON 2700
2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGTPVSDSVSEKAPNIIKSKON 2700
2701 QAKQNVGNGVPMRTVGLNENLNFIOVDAPDQGTIKQGNPNPVSETNNESSIIVERT 2760
2701 QAKQNVGNGVPMRTVGLNENLNFIOVDAPDQGTIKQGNPNPVSETNNESSIIVERT 2760
2761 PFSSSSSSKHSSPGSTVAARVTPFNYPNPSPRKSSADSTTSARPSQIPTPVNNTKRDST 2820
2761 PFSSSSSSKHSSPGSTVAARVTPFNYPNPSPRKSSADSTTSARPSQIPTPVNNTKRDST 2820
2821 DSTESSGTSQPKRHSGSYLVTSV 2843
2821 DSTESSGTSQPKRHSGSYLVTSV 2843

RESULT 6
AAV72782 standard; protein; 2973 AA.
XX AAV72782;
XX AC AAV72782;
XX DT 31-MAY-2001 (first entry)
XX DE Transcriptional activation protein #1 related to the invention.
XX KW Human; Tcf-4 protein; transcriptional activation factor; beta-catenin;
XX KW apoptosis; colorectal cancer; Familial Adenomatous Polyposis; FAP; APC;
XX KW adenomatous polyposis coli; recombinant adenovirus; Ad-Mini-Me; therapy;
XX KW GFP; green fluorescent protein; GFP/cAPC fusion protein; cytostatic.
XX OS Homo sapiens.
XX PN WO200116167-A2.
XX PD 08-MAR-2001.
XX PF 29-AUG-2000; 2000WO-US023635.
XX PR 01-SEP-1999; 99US-00388354.
XX PA (UYJO) UNIV JOHNS HOPKINS.
XX PI Barker N, Clevers JC, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
XX PI Vogelstein B, He T;
XX WI WIPI; 2001-226675/23.
XX PT New fusion protein comprising an enzyme covalently linked to a portion of
XX PT the adenomatous polyposis coli comprising its beta-catenin binding domain
XX PT useful for inducing apoptosis or treating colorectal cancer.
XX FS Disclosure; Page 74-81; 83pp; English.
XX CC The present sequence is a transcriptional activation protein related to
XX CC the invention. The invention relates to human Tcf-4 proteins and their
XX CC corresponding cDNA molecules which encodes transcriptional activation
XX CC factors of human Tcf/Lef family. Human Tcf-4 binds to beta-catenin and
XX CC activates transcription in colorectal epithelial cells. Moreover it has
XX CC been found that adenomatous polyposis coli (APC) regulates this
XX CC transcriptional activation, by binding to beta-catenin. The invention
XX CC also provides a recombinant adenovirus, Ad-Mini-Me i.e., APC Minus its
XX CC amino- and carboxyl-terminal Ends which expresses a fusion protein,
XX CC green fluorescent protein (GFP)/cAPC containing GFP fused to the central
XX CC third of APC which contains its beta-catenin binding domain useful for
XX CC inducing apoptosis or treating colorectal cancer. These fusion proteins
XX CC are useful for treating cancer, e.g. colorectal cancer, and other cancers
XX CC associated with Familial Adenomatous Polyposis (FAP) or patients with
XX CC increased risk of developing cancer. Human Tcf-4 cDNA provides an
XX CC excellent system for screening agents for their ability to promote
XX CC delivery, integration, hybridization, expression, replication or
XX CC integration in cells or in an animal. It also provides methods for
XX CC diagnosis cancer in a sample suspected of being neoplastic
XX SQ Sequence 2973 AA;

Query Match 100.0%; Score 14566; DB 4; Length 2973;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAASVDQLLKQVEALKMNSNLRLQELDNNSHLTKLETSANMKVELKQLQSGSIEDAM 60
Db 1 MAAASVDQLLKQVEALKMNSNLRLQELDNNSHLTKLETSANMKVELKQLQSGSIEDAM 60
QY 61 ASSGQIDLLERLKLNLDSNFPVGLRSKMSLSRYSGRSGSVSSRSGECSPPVPMGSPFR 120
Db 61 ASSGQIDLLERLKLNLDSNFPVGLRSKMSLSRYSGRSGSVSSRSGECSPPVPMGSPFR 120
QY 121 RGFVNGRSGSTGYLELEKERSLLADLDKEKEKDWYAAQLNLTKRIDSJLPTENFSL 180

DB 121 RGFVNSRSTGVLEBLEKERSULLADLDKEKEDWYIAQONLTKRIDSPLTFNPSL 180
QY 181 QTDMTROLEYEARQIRVAMEEQBLGTCODMEKRAQRRIARIQOIEBKDILIRIQLLOSOAT 240
DB 181 QTDMTROLEYEARQIRVAMEEQBLGTCODMEKRAQRRIARIQOIEBKDILIRIQLLOSOAT 240
QY 241 EAERSSONKHETGSHDAERONEGQVGEINMATSGNGQSTTRMDHETASVLSSSSTHSA 300
DB 241 EAERSSONKHETGSHDAERONEGQVGEINMATSGNGQSTTRMDHETASVLSSSSTHSA 300
QY 301 PRRLTSHLGTKVEMYVLSLLSMGLGTHDKDDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
DB 301 PRRLTSHLGTKVEMYVLSLLSMGLGTHDKDDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
QY 361 HGNDKDSVLLGNRSRGSKEARASAAALHNIHSPQDDKRGREIRVHLHBOIRAYCETC 420
DB 361 HGNDKDSVLLGNRSRGSKEARASAAALHNIHSPQDDKRGREIRVHLHBOIRAYCETC 420
QY 421 WEMQEAHEPGMDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGIQAIAELLQ 480
DB 421 WEMQEAHEPGMDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGIQAIAELLQ 480
QY 481 VDCEMYGLTNDHYSITLRIYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSESEDL 540
DB 481 VDCEMYGLTNDHYSITLRIYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSESEDL 540
QY 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSVLNALWLSAHT 600
DB 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSVLNALWLSAHT 600
QY 601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILRVNSSLIATNEDHROILRENN 660
DB 601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILRVNSSLIATNEDHROILRENN 660
QY 661 CLQTLLOHLKSHSLTVSNACOTLWNLARPQDOEALWDMGAVSMLXNLHSHKHMIA 720
DB 661 CLQTLLOHLKSHSLTVSNACOTLWNLARPQDOEALWDMGAVSMLXNLHSHKHMIA 720
QY 721 GSAALRNLMANRPAYKIDANIMSPGSSLP SLHVRKQKALAEALDAOHLSTTFNIDNLS 780
DB 721 GSAALRNLMANRPAYKIDANIMSPGSSLP SLHVRKQKALAEALDAOHLSTTFNIDNLS 780
QY 781 PKASHRSQRKHQSILYGYVFTDNRHDDNRDNFNTGNTVLSPLYNTTVLPSSSSSRGS 840
DB 781 PKASHRSQRKHQSILYGYVFTDNRHDDNRDNFNTGNTVLSPLYNTTVLPSSSSSRGS 840
QY 841 LDSSRSKDRSLERERGLGNYPHATENPGTSSKRGLOISTTAAQIAKMWEEVSAIHTS 900
DB 841 LDSSRSKDRSLERERGLGNYPHATENPGTSSKRGLOISTTAAQIAKMWEEVSAIHTS 900
QY 901 QEDRSSGTTTELHCVTDERNALRRSSAAHTSNFTYNTFKSENSNRTCSMPYAKLEYKRSS 960
DB 901 QEDRSSGTTTELHCVTDERNALRRSSAAHTSNFTYNTFKSENSNRTCSMPYAKLEYKRSS 960
QY 961 NDSLNSVSSDGYKRGOMKPSIESYSEDDSKFCSYQYPADLAHKHSHANWDDNDGE 1020
DB 961 NDSLNSVSSDGYKRGOMKPSIESYSEDDSKFCSYQYPADLAHKHSHANWDDNDGE 1020
QY 1021 LDTPTINYSLKYSDEQLNSGRQSPONERWARPKHIIIEDEIKQSEQRORSNOSTTYPVYTE 1080
DB 1021 LDTPTINYSLKYSDEQLNSGRQSPONERWARPKHIIIEDEIKQSEQRORSNOSTTYPVYTE 1080
QY 1081 STDDKHLXFPQHFQGOECVSPYRGRGANGSTNRVGSNHGINQVNSQSLQOEDDYEDDKP 1140
DB 1081 STDDKHLXFPQHFQGOECVSPYRGRGANGSTNRVGSNHGINQVNSQSLQOEDDYEDDKP 1140
QY 1141 TNYSERYSEREEHREERTNYSIKYNEEKHVDQPIDYSLKYATDTPSSOKQSFPSKS 1200
DB 1141 TNYSERYSEREEHREERTNYSIKYNEEKHVDQPIDYSLKYATDTPSSOKQSFPSKS 1200
QY 1201 SSGQSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRSQPOKAATCKVSSINQETIQ 1260

DB 1201 SSGQSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRSQPOKAATCKVSSINQETIQ 1260
QY 1261 TYCVEDTPICFSRCSLSLSAASAEIGCNOTTQCEADSAANTLOIAEIKKEKIGTRSADPV 1320
DB 1261 TYCVEDTPICFSRCSLSLSAASAEIGCNOTTQCEADSAANTLOIAEIKKEKIGTRSADPV 1320
QY 1321 SEVPAVQCHPRTKSRRLQGSLSLSSESARHKAVFSSGAKSPSKSGAQTPKSPPEHYVOET 1380
DB 1321 SEVPAVQCHPRTKSRRLQGSLSLSSESARHKAVFSSGAKSPSKSGAQTPKSPPEHYVOET 1380
QY 1381 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSPCOTMPPSRSTTP 1440
DB 1381 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSPCOTMPPSRSTTP 1440
QY 1441 PPPOTAQTKREVPNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLLHFAESTPDGF 1500
DB 1441 PPPOTAQTKREVPNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLLHFAESTPDGF 1500
QY 1501 SCSSLSALSILDEPFFIQKDVLEIRIMPPVQENDNGNETESEPQKESNENQKEAKTIDSE 1560
DB 1501 SCSSLSALSILDEPFFIQKDVLEIRIMPPVQENDNGNETESEPQKESNENQKEAKTIDSE 1560
QY 1561 KDLLDDDDDDIIELEECIIISAMPTKSSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPS 1620
DB 1561 KDLLDDDDDDIIELEECIIISAMPTKSSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPS 1620
QY 1621 QNRLOPKHYSVTFPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGSVRGAQ 1680
DB 1621 QNRLOPKHYSVTFPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGSVRGAQ 1680
QY 1681 SGFEFEKEDTITPTEGRSTDEAOGGKTSSVTIPELDDNKAKEGDIILAEICINSAMPKSHKP 1740
DB 1681 SGFEFEKEDTITPTEGRSTDEAOGGKTSSVTIPELDDNKAKEGDIILAEICINSAMPKSHKP 1740
QY 1741 FRVKKIMDVQOQASASSAPNKNQLOGKKKKPTSPVKPIQNTTEYTRVRKNADSKNNLN 1800
DB 1741 FRVKKIMDVQOQASASSAPNKNQLOGKKKKPTSPVKPIQNTTEYTRVRKNADSKNNLN 1800
QY 1801 AERFSDNKKONLKNNSKDFNDKLPNNEDVRGSFAFDSPHHYTPIEGTFCYCSRND 1860
DB 1801 AERFSDNKKONLKNNSKDFNDKLPNNEDVRGSFAFDSPHHYTPIEGTFCYCSRND 1860
QY 1861 SLSLDPDDDDVLSREKAEFLRKAKENKSEAKVTSHTELTSNQQSANKTOAIKOPINR 1920
DB 1861 SLSLDPDDDDVLSREKAEFLRKAKENKSEAKVTSHTELTSNQQSANKTOAIKOPINR 1920
QY 1921 GQPKPILQKOSTTPOSSKOIPDRGAATDEKLQNFATENTPVCFSHNSLSLSIDIOENN 1980
DB 1921 GQPKPILQKOSTTPOSSKOIPDRGAATDEKLQNFATENTPVCFSHNSLSLSIDIOENN 1980
QY 1981 NKENEPIKETEPPDSQGEPSKPOASGYAPKSFHVEDTPVCFSENSSLSIDSEDDLQ 2040
DB 1981 NKENEPIKETEPPDSQGEPSKPOASGYAPKSFHVEDTPVCFSENSSLSIDSEDDLQ 2040
QY 2041 ECISAMPKPKKPSRLKGDNEKHSRNMGGIIGEDJTLDLKOIQRDPDSEHGLSPDSENF 2100
DB 2041 ECISAMPKPKKPSRLKGDNEKHSRNMGGIIGEDJTLDLKOIQRDPDSEHGLSPDSENF 2100
QY 2101 WKAIQSGANSIVSSLHQAACCLSPQASDSDSLSLKSGISLQSGPHLTPDOEEKPFT 2160
DB 2101 WKAIQSGANSIVSSLHQAACCLSPQASDSDSLSLKSGISLQSGPHLTPDOEEKPFT 2160
QY 2161 SNKGPRILKPGEKSTLTETKKIESEKGIKGGKKVYKSLITGKVRNSNSEISGQWKQPLQAN 2220
DB 2161 SNKGPRILKPGEKSTLTETKKIESEKGIKGGKKVYKSLITGKVRNSNSEISGQWKQPLQAN 2220
QY 2221 MPSISRGRTMIHLPVGRNSSSSTSPVSKKGPPLKTPPASKSPSEGOATTATTPRGAKPSVK 2280
DB 2221 MPSISRGRTMIHLPVGRNSSSSTSPVSKKGPPLKTPPASKSPSEGOATTATTPRGAKPSVK 2280
QY 2281 ELSPVARQTSQIGGSSKAPSRSGRSDTSPRPAQQLSRPIQSPGRNSISPGRNGISPPN 2340
DB 2281 ELSPVARQTSQIGGSSKAPSRSGRSDTSPRPAQQLSRPIQSPGRNSISPGRNGISPPN 2340

